PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶:
G06F 17/50, 159/00, G06G 7/58, G01N
24/12, 33/53

(11) International Publication Number:

WO 96/30849

(43) International Publication Date:

3 October 1996 (03.10.96)

(21) International Application Number:

PCT/US96/04229

A1

(22) International Filing Date:

27 March 1996 (27.03.96)

(30) Priority Data:

418,992

31 March 1995 (31.03.95) U

us Ì

(71) Applicant: CURAGEN CORPORATION [US/US]; 322 East Main Street, Branford, CT 06405 (US).

(72) Inventors: DEEM, Michael, W.; 24 Highland Avenue #14, Cambridge, MA 02139 (US). ROTHBERG, Jonathan, M.; 49 Old Quarry Road, Guilford, CT 06437 (US). WENT, Gregory, T.; 34 Scotland Avenue, Madison, CT 06443 (US).

(74) Agents: MORRIS, Francis, E. et al.; Pennie & Edmonds, 1155 Avenue of the Americas, New York, NY 10036 (US). (81) Designated States: AL, AM, AU, AZ, BB, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, IS, IP, KG, KP, KR, KZ, LK, LR, LS, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TR, TT, UA, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: CONSENSUS CONFIGURATIONAL BIAS MONTE CARLO METHOD AND SYSTEM FOR PHARMACOPHORE STRUCTURE DETERMINATION

(57) Abstract

In a specific embodiment, this invention comprises a method for selecting highly targeted lead compounds for design of a drug that binds to a target molecule. The method comprises screening a diversity library against the target molecule of interest to pick the selectively binding members. Next the structure of the selected members is examined and a candidate pharmacophore responsible for the binding to the target molecule is determined. Next, preferably by REDOR nuclear magnetic resonance, several highly accurate interatomic distances are determined in certain of the selected members which are related to the candidate pharmacophore. A highly accurate consensus, configurational bias, Monte Carlo method determination of the structure of the candidate pharmacophore is made using the structure of the selected members and incorporating as constraints the shared selected members and incorporating as constraints the shared candidate pharmacophore and the several measured distances. This determination is adapted to efficiently examine only relatively low energy configurations while respecting any structural constraints present in the organic diversity library. If the diversity library contains short peptides, the determination respects the known degrees of freedom of peptides as well as any internal constraints, such as those imposed by disulfide bridges. Finally, the highly accurate pharmacophore so determined is used to select lead organics for drug development targeted at the initial target molecule.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

		GB	United Kingdom	MW	Malawi
AM	Armenia	GE	_	MX	Mexico
AT	Austria		Georgia	NE.	
AU	Australia	GN	Guines		Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	italy	PL	Poland
BJ	Benin	JP	Japan -	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic	SD	Sudan
CF	Central African Republic		of Korea	SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	u	Liechtenstein	SK	Slovakia
СМ	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
cs	Czechoslovakia	LT	Lithumia	TD	Chad
CZ	Czech Republic	LU	Luxenbourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco '	11	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FT	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

CONSENSUS CONFIGURATIONAL BIAS MONTE CARLO METHOD AND SYSTEM FOR PHARMACOPHORE STRUCTURE DETERMINATION

This specification includes in Sec. 8 computer program

5 listings that are exemplary embodiments of the computer programs of this invention.

A portion of the disclosure of this patent document contains material which is subject to copyright protection.

The copyright owner has no objection to the facsimile reproduction by any one of the patent disclosure, as it appears in the Patent and Trademark Office patent files and records, but otherwise reserves all copyright rights whatsoever.

This invention was made with Government support under

Grant number 1R43CA62752-01 awarded by the National
Institutes of Health. The Government has certain rights in the invention.

1. FIELD OF THE INVENTION

The field of this invention is computer assisted methods of drug design. More particularly the field of this invention is computer implemented smart Monte Carlo methods which utilize NMR and binders to a target of interest as inputs to determine highly accurate molecular structures that must be possessed by a drug in order to achieve an effect of interest. Illustrative U.S. Patents are 5,331,573 to Balaji et al., 5,307,287 to Cramer, III et al., 5,241,470 to Lee at al., and 5,265,030 to Skolnick et al.

30

2. BACKGROUND

Protein interactions have recently emerged as a fundamental target for pharmacological intervention. For example, the top two major uncured diseases in the United States are atherosclerosis (the principal cause of heart attack and stroke) and cancer. These diseases are

responsible for greater than 50% of all U.S. mortality and cost the U.S. economy over \$200 billion per year. A consistent picture of these diseases, which has gradually emerged during the past ten years of molecular biological and 5 medical research, views both as triggered by disordering of specific molecular recognition events that take place among sets of proteins present in both the normal and disease states.

Hierarchical, organized patterns of protein-protein

10 interactions are often referred to as "pathways" or

"cascades." At the molecular level, cancers have been
determined to be the deregulation of pathways of interacting
proteins responsible for guiding cellular growth and
differentiation. During the past year, individual cellular

15 events have been organized into nearly complete mechanistic
explanations of how a cell's behavior is controlled by its
environment and how communication pathway errors lead to
uncontrolled proliferation and cancer. Disruption in similar
pathways are responsible for the proliferation of blood

20 vessel walls marking the atherosclerotic disease state (Cook
et al., 1994, Nature 369:361-362; Hall, 1994, Science
264:1413-1414; Ross, 1993, Nature 362:801-809; Zhang et al.,
1993, Nature 364:308-313).

Inhibition or stimulation of particular protein25 substrate interactions have long been known drug targets.

Many important anti-hypertensives, neurotransmitter

analogues, antibiotics, and chemotherapeutic agents act in
this fashion. Captopril, an antihypertensive drug, was
designed based on its ability to antagonize a focal blood30 pressure-regulating enzyme.

proteins involved in biological processes, either as part of protein-protein pathways or as enzymes, are composed of domains (Campbell et al., 1994, Trend. BioTech. 12:168-172; Rothberg et al., 1992, J. Mol. Biol.

35 227:367-370). Domains, or regions of the protein of stable three dimensional (secondary and tertiary) structures, play several major roles, including providing on their surface

small regions ("examples of targets"), where proteins and substrates are able to bind and interact, and functioning as structural units holding other domains together as part of a large protein (tertiary and quaternary structure). The interaction surface of a domain or target is fundamental to determining binding specificity. Targets are often small enough that the principal contribution to the binding energy is short range, highly localized to several amino acids (Wells, 1994, Curr. Op. Cell Biol. 6:163~174). The

- 10 functional specificity of targets and domains, responsible for the incredible diversity of cellular function, ultimately rests with the arrangement of amino acid side chains forming their interaction surfaces, or targets (Marengere et al., 1994, Nature 369:502-505).
- 15 It can be appreciated, therefore, that pharmacological intervention affecting the specific protein-protein and protein-substrate recognition events occurring at protein targets is of fundamental importance, particularly for effective drug design.
- However, achieving desired pharmacological interventions in a predictable manner remains as elusive as ever. Early approaches to drug design depended on the chance observation of biological effects of a known compound or the screening of large numbers of exotic compounds, usually derived from 25 natural sources, for any biological effects. The nature of the actual protein target was usually unknown.

2.1. TARGET STRUCTURE-BASED APPROACHES TO DRUG DESIGN

Rational approaches to drug design have met with only limited success. Current rational approaches are based on first determining the entire structure of the proteins involved in particular interactions, examining this structure for the possible targets, and then predicting possible drug molecules likely to bind to the possible target. Thus the location of each of the thousands of atoms in a protein must be accurately determined before drug design can begin.

Direct experimental and indirect computational methods for protein structure determination are in current use. However, none of these methods appears to be sufficiently accurate for drug design purposes according to current rational.

5 approaches.

The primary direct experimental methods for determining the structure of proteins involved in particular interactions are X-ray crystallography, relying on the interaction of electron clouds with X-rays, and liquid nuclear magnetic 10 resonance (NMR), relying on correlations between polarized nuclear spins interacting via indirect dipole-dipole interactions. X-ray methods provide information on the location of every heavy atom in a crystal of interest accurate to 0.5-2.0 Å (1 Å = 10^{-8} cm). Drawbacks of x-ray 15 methods include difficulties in obtaining high-quality crystals, expense and time associated with the crystallization process, and difficulties in resolving whether or not the structure of the crystalline forms is representative of the in vivo conformation (Clore et al., 20 1991, J. Mol. Biol. 221:47; Shaanan et al., 1992, Science 227:961-964). High resolution, multidimensional, liquid phase NMR techniques represent an attractive alternative, to the extent that they can be applied in situ (i.e., in aqueous environment) to the study of small protein domains (Yu et 25 al., 1994, Cell 76:933-945). However, the complexity of the analysis of the various mutual correlations is time consuming, and the correlations (primarily from the nuclear Overhausser effect) provide no better accuracy than X-ray methods. Isotopic enrichment of proteins with 13C and 15N 30 reduces the time associated with analysis, but at a great expense (Anglister et al., 1993, Frontiers of NMR in Biology III LZ011). ,

Protein structures determined by any of these current methods do not predict success in subsequent drug design.

35 Resolution obtainable either by measurement or computation, generally 0.5-2 Å, has often been found to be inadequate for effective direct drug design, or for selection of a lead

compound from organic compound libraries. The resolution required to understand both drug affinity and drug specificity, although not precisely known, is probably measured in fractions of an Å, down to 0.1 Å (MacArthur et al., 1994, Trend. BioTech. 12:149-153). This accuracy appears to be beyond the capabilities of many current methodologies.

Prior research has identified tools which, although promising, cannot be used in a coordinated manner for drug 10 design. One promising measurement approach with speed, simplicity, accuracy, and the ability to carefully control the measurement environment is rotational echo double resonance (REDOR) NMR, a type of solid state NMR (Guillion and Schaefer, 1989, J. Magnetic Resonance 81:196; Holl et 15 al., 1990, J. Magnetic Resonance 81:620-626 and McWherter, 1993, J. Am. Chem. Soc. 115:238-244). REDOR accuracy can be below the 0.1 Å believed to be sufficient for direct drug design. However, since REDOR measures only a few selected distances, it is not usable in drug design methods which 20 depend on the initial determination of the complete structure of the protein containing the target of interest.

Once a target's structure is determined by the above methods, most rational drug design paradigms call for the prediction of small drug structures that will bind (or dock) 25 to the target. This prediction is generally done by computational methods, of which several are in current use. Most seek to predict the position of all the thousands of atoms in a drug structure. Purely ab initio computational approaches to high resolution structure analysis, such as 30 quantum statistical mechanics and molecular dynamics, require prohibitive computing resources. To apply either approach, the potential energy, or Hamiltonian, of the entire system must be known. Statistical mechanics provides an expression for the probability of any given protein configuration as a 35 ratio of partition functions. Proper quantum statistical mechanics required for an exact evaluation of full protein partition functions is not currently computationally

feasible, as it would involve many thousands of atoms including the target, the protein, and the aqueous environment. The application of even simple, approximate quantum statistical mechanics to simple systems in aqueous 5 environments is currently a non-trivial task (Chandler, 1991, in Liquids, Freezing, and Glass Transitions, Elsevier, NY, p. Molecular dynamics computes the dynamics of a molecule's motion in time. Computing the atomic dynamics of all the perhaps thousands atoms of a protein is an extreme 10 computational burden. Only picoseconds, or at most a few nanoseconds, of molecular time can be simulated, which is insufficient to determine a high resolution, equilibrium, structure (Smit et al., 1994, J. Phys. Chem. 98:8442-8452). In any case, most of the information determined is wasted, 15 since only the structure of the protein binding target are of interest in drug design.

Further, current approximate computational techniques for protein structure determination are in need of greater accuracy or efficiency. The most common techniques depend on 20 Molecular Dynamics or Monte Carlo methods (Nikiforovich, 1994, Int. J. Peptide Protein Res. 44:513-531; Brünger and Karplus, 1991, Acc. Chem. Res. 24:54-61). These methods randomly alter initial molecular structures by generating simulated thermal perturbations, and then average the 25 ensemble of results to determine a final structure. generated perturbation must preserve all structural constraints and be energetically favorable. If both conditions are not met, the perturbation will be discarded. Current Monte Carlo methods applied to constrained protein 30 structure determinations productively use only approximately 1 out of 105 perturbed structures generated (Siepmann et al., 1993, Nature 365:330-332). This extreme waste of computer

35 To summarize, existing rational drug design methods based on identification of target structure fail to reliably yield drug molecules due to experimental structure

determinations.

rescurces results in time consuming, low resolution structure

determination difficulties and computational difficulties associated with predicting drug structures with ill-defined Hamiltonians.

2.2. DIVERSITY-BASED APPROACHES TO DRUG DESIGN

Another method for exploring protein target interactions utilizes "recognition systems" which comprise huge libraries of related molecules (Clarkson et al., 1994, Trend. BioTech. 12:173-184). From such a library only those members binding 10 to the target of interest are selected. Such recognition systems must encompass the structural diversity of protein targets while being amenable to serve for the selection of lead compounds for drug design. Antibodies are one classic example of such a system that certainly meets the recognition 15 requirement. Unfortunately, there is a need to determine the antibody structures needed for lead compound selection more rapidly and accurately. While about 2000 recognition regions have been sequenced, only about 23 in the Brookhaven Protein Structural Database have structures determined to even within 20 2 Å (Rees et al., 1994, Trends in Biotech. 12:199-206).

Promising recognition systems at the opposite extreme comprise huge libraries of small peptides. The small peptides must be sufficiently diverse so that they attain a level of affinity and specificity similar to that obtained by 25 protein domains. Given the role peptides play in nature, this condition can be met by surprisingly small structures, with 6 to 12 amino acids. However, linear peptides are either unstructured or weakly structured at room temperature in aqueous solutions (Alberg et al., 1993, Science 262:248; 30 Skalicky et al., 1993, Protein Science 10:1591-1603). From a practical viewpoint, linear peptides must be constrained to reduce their degrees of freedom (reduced conformational entropy) and to increase their chances for strongly binding. These constraints, or scaffolds, limit the range of stable 35 conformations and make more straightforward determining bound structure (Olivera et al., 1990, Science 249:259; Tidor et

al., 1993, Proteins: Structure Function and Genetics 15:71).

Methods are now available to create such libraries and to select library members that recognize a specific protein target. The production of constrained peptide diversity libraries requires synthesizing oligonucleotides with the 5 desired degeneracy to code for the peptides and ligating them into selection vectors (Goldman et al., 1994, Bio/Tech. 10:1557-1561). Once a constrained structured diversity library is created, it is a source from which to select specific members that bind to a target of interest. Beginning 10 with a known pathway involving specific domain-domain or protein-substrate interactions at a target, molecular biological methods can be used to identify in a matter of days small ensembles of highly constrained peptides from these huge libraries that bind to these domains with high 15 affinity and specificity.

While this field has been exploding in the last few years and showing great potential, it is severely limited by its use in isolation without the benefit of integrated structural analysis needed both to derive the high resolution 20 structures of binding peptides and also to direct the construction of additional structured libraries. Drug design is not aided by having library members recognizing the protein target of interest but without any understanding of why the recognition occurs. This is entirely similar to the 25 random screening methods of early fortuitous drug design efforts.

Unfortunately, rational drug design according to current approaches (target structure-based) remains an inefficient, laborious process with a disproportionately high lead30 compound failure rate. Presently, about 90% of lead compounds fail to emerge successfully from clinical trials (Trends in U.S. Pharmaceutical Sales and Research and Development, Pharmaceutical Manufacturing Association, Washington, D.C., 1993).

It is becoming clear that low-resolution structures of an entire protein or target (at 0.5-2 \mathring{A}), or an

uncharacterized lead, such as produced by chemical diversity methods, leave much to be desired for use in drug design.

If the limitations of prior art methods were overcome and a sufficiently accurate structure needed by a molecule to 5 bind to a target of interest could be determined, existing chemical libraries could be searched for highly targeted lead compounds with similar structure (Martin, 1992, J. Medicinal Chem. 35:2145-2154). This database search can be based not only on chemical and electronic properties, but also on 10 geometric information. Such searches that have high resolution (better than 0.25 Å), would provide a vast improvement over the prior art, as lower resolutions lead to an exponentially increasing number of potential leads.

Computational methods to determine high resolution drug
15 structures from recognition system binding information or NMR
partial distance measurements are not currently available.
No current structure determination methods uses such
additional information to make more efficient or more
accurate determination of high resolution structures
20 (Holzman, 1994, Amer. Sci. 872:267).

Citation of a reference or discussion hereinabove shall not be construed as an admission that such is prior art to the present invention.

3. SUMMARY OF THE INVENTION

It is a broad object of this invention to address the prior art problems of drug design by providing a method of rational design of drugs that achieve their effect by binding to a target molecule or molecular complex of interest.

- 30 Importantly, this object is achieved without requiring determination of the structure of the molecule or molecular complex ("target molecule") bearing the target or even of the target itself. The method is target structure independent. The method of the invention uses an interdisciplinary
- 35 combination of computational modeling and simulation, experimental distance constraints, and molecular biology.

In an important aspect, the invention provides a computer implemented modeling and simulation method to determine a highly accurate consensus structure for the pharmacophore and a structure for the remainder of the 5 molecule from diversity library members that bind to the protein target of interest. Where prior structure determination methods focused on the structure of the target molecule or of the target, the method of this invention is uniquely adapted to focus instead on the structures of 10 molecules that bind to the target. Such structural information is directly applicable to drug design since it defines the structure a drug must possess to bind to the target of interest. Also, this structural information is much easier to determine by use of the present invention, 15 since it concerns molecules with many fewer atoms than the target molecule. The method of the invention achieves accuracy by improving upon the accuracy and utility of the input structural information. In a further embodiment of the invention, the method employed for structural determination 20 is a smart Monte Carlo technique adapted to small constrained molecules.

The structure determination method of the invention allows one to take maximum advantage of the information obtained from the molecular biological selection of the 25 diversity library members that tightly and specifically bind to the target molecule of interest. The selected library members must share some common structure to bind to the same target molecule. The smart Monte Carlo computer method of this invention specifically seeks and provides this common structure.

The invention also provides a method of performing REDOR NMR measurements of molecules on a solid phase substrate. In a preferred embodiment, the substrate is a solid phase on which the molecule (e.g., peptide) has been synthesized, with 35 a high degree of purity. In another preferred embodiment, performing REDOR measurements of such a molecule on a substrate can be done in a dry nitrogen atmosphere, under

hydrated conditions, and when the molecule is either free or bound to a target. In a specific embodiment, the REDOR measurements are accurate to better than 0.05 Å from 0 to 4 Å, and to better than 0.1 Å from 4 to 8 Å. In an 5 advantageous aspect of the invention, the structure determination method makes maximum use of these highly accurate internuclear distance measurements to constrain the determined common structure for the binding library members.

The invention also provides methods of identifying a 10 compound that specifically binds to a target molecule, by first screening a diversity library, and then using a genetic selection method for screening the compounds identified from the diversity library.

In broad aspects, the invention provides a method and
15 apparatus for rational and predictable design of new and/or
improved drugs that achieve their effect by binding to a
specified target molecule. More particularly, the invention
is directed to a method for the rational selection of highly
specific lead compounds for such drug design, including the
20 computer implemented step of highly accurate determination of
the structure responsible for this target binding by the
highly accurate, consensus, configurational bias Monte Carlo
method.

A lead compound serves as a starting point for drug

25 development both because it specifically binds to the protein target of interest, achieving the biological effect of interest, and because it has or can be modified to have good pharmacokinetics and medicinal applicability. A final drug may be the lead compound or may be derived therefrom by

30 modifying the lead to maximize beneficial effects and minimize harmful side-effects. Although any lead compound is useful, a lead that tightly and specifically binds to the target molecule of interest in a known manner, such as can be provided by the invention, is of great use. Knowledge of the

35 high resolution structures in a lead compound responsible for its binding and activity provides a more focused and efficient drug development process.

The methods of the invention improve lead compound determination, by determining the "pharmacophore", the precise structural characteristics needed for a lead compound to specifically bind to a target of interest. The most 5 fundamental specification of a pharmacophore is in terms of the electronic properties necessary for a molecule to specifically bind to the surface of a target molecule. These properties may be fundamentally represented by requirements on the ground and low lying excited state wave functions of a 10 pharmacophore, such as, for example, by specifying requirements on the well known multiple expansion of these wave functions.

The preferred pharmacophore specification according to the invention is in terms of both the chemical groups making 15 up the pharmacophore and determining its electronic properties and also the geometric relationships of these groups. This chemical representation is not the only possible representation of the pharmacophore. Several chemical arrangements may have similar electronic properties.

20 For example, if a pharmacophore specification included an -OH group at a particular position, a substantially equivalent specification might include an -SH group at the same position. Equivalent chemical groups that may be substituted in a pharmacophore specification without substantially 25 changing its nature are called "homologous".

In particular embodiments, therefore, this invention provides a method and apparatus for the highly accurate determination of the pharmacophore needed to specifically bind to the target molecule of interest, by a specification 30 of the geometric relationships of the important chemical groups. The pharmacophore is preferably determined by a smart Monte Carlo method from molecular biological input specifying molecules (preferably selected from among diversity libraries) that specifically bind to the target 35 molecule and also preferably from REDOR NMR data specifying a few highly accurate distances in these selected molecules.

An important advantage provided by the invention is the ability to make a pharmacophore structure determination without relying on any knowledge of the structure of the target molecule or target. Where the target molecule is a 5 protein, conventional prior art methods have sought to sequence and determine the structure of the protein containing the target, hoping thereby to determine active sites by examination of the structure. A further important advantage of the invention is that this structure

- 10 determination can be made by use of a relatively small number of actual physical position measurements. In contrast, conventional methods using X-ray crystallography and liquid NMR require determination of positions of all atoms in the molecule ("binder") that specifically binds to the target,
- 15 and the target. An additional advantage provided by the invention is that, in a preferred embodiment wherein REDOR structural measurements provide input information, the accuracy of the pharmacophore structure determination can be at least approximately 0.25-0.50 Å or better. This accuracy
- 20 is provided by the combination of an efficient, Monte Carlo technique for structure determination with a few highly accurate distance determinations.

4. BRIEF DESCRIPTION OF THE DRAWINGS

- These and other features, aspects, and advantages of the present invention will become better understood by reference to the accompanying drawings, following description, and appended claims, where:
- Fig. 1 is the overall method of this invention in its 30 broadest aspect;
 - Fig. 2A and 2B are more detail for the step of Fig. 1 for selecting candidate pharmacophore structures;
 - Fig. 3 is more detail for the step of Fig. 1 for preforming distance measurements;
- Fig. 4 is more detail for the step of Fig. 3 for performing NMR measurements;

Fig. 5 is REDOR NMR signal response details for step of Fig. 3 of data analysis;

- Fig. 6 is sample REDOR NMR spectra according to the method of Fig. 3;
- Fig. 7 is sample data analysis according to the method of Fig. 3;
 - Fig. 8 is more detail for the Step of Fig. 1 for configurational bias Monte Carlo structure determination;
 - Fig. 9 is a sample of simulation completion data;
- Fig. 10 is further detail of peptide memory representation used in the method of Fig. 8;
 - Fig. 11 is additional detail of peptide memory representation used in the method of Fig. 8;
- Fig. 12 is more detail for the step of Fig. 8 of
 15 processor generation of proposed modified structures by Type
 I moves:
 - Fig. 13 is more detail for the step of Fig. 8 of processor generation of proposed modified structures by Type II moves:
- 20 Fig. 14 is additional detail for the step of Fig. 8 of processor generation of proposed modified structures by Type II moves;
 - Fig. 15 is a structure for implementing the method of Fig. 8;
- Fig. 16 is the main program structure of Fig. 15;

 Fig. 17 is the structure modification program structure of Fig. 15;
 - Fig. 18A and 18B are the Type I move generator program structure of Fig. 17;
- Fig. 19A and 19B are the Type II move generator program structure of Fig. 17.

5. DETAILED DESCRIPTION

For clarity of disclosure, and not by way of limitation, 35 the detailed description of the invention is described as a series of steps. A broad view of the exemplary steps of which the invention is comprised is presented in Fig. 1, a

```
brief overview of which is presented in the text that
                                                                                                                                                                                                            The invention method preferably begins with a target range for molecular complex)
                                                                                                                                            The invention method preferably begins with a target of specific binding of molecular complex) interest.

The invention method preferably begins with a target of target of binding of specific binding binding binding of specific binding binding binding binding binding of specific binding specific binding of specific binding binding binding of specific binding bindi
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                specific binding of
                                                                                                                                                                               biological or pharmacological interest. specific bir affect its affect affect is predicted to affect affect affect is predicted hindrainal affect and may provide hindrainal affect amolecule to the target and may provide hindrainal affect aff
                                                                                                                                                                                         a molecule to the target may provide biological include the target and these effects might include biological for example.
                                                                                                                                   molecule (or molecular complex) interest.

5 biological or pharmacological interest.
WO 96/30849
                                                                                                                                                                                               amelioration of a disease process or alteration of a the holecule amelioration of a disease Lead compounds to target molecule amelioration response. Specifically bind to target molecule invention are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule amelioration are able to specifically bind to target molecule and target molecule amelioration are also and target molecule ame
                                                                                                                                                                                                                          interest. For example, these effects might include a disease process or alteration of a anelioration o
                                                                                                                                                                                                                                                             physiological response. specifically bind to target molecule 1

physiological able to specifically points for the design of a drug points for the drug points 
                                                                                                                              follows.
                                                                                                                                                                                                                                                                             invention are able to specifically bind to target molecule to specifically bind to the design of a drug and can specifically bind to the target.
                                                                                                                                                                                                                                                                                                                                                                                   to the target. allows the to specifically bind to the target. Prom among library mambare of a circulation from 
                                                                                                                                                                                                                                                                       Diversity library screening, step 2, allows the of a plurality of members of a plurality of rhar snerification from among library members "library members" rhar snerification from among called "binders" har snerification inereinafter called "binders" plurality of allows the plurality of allows the property of allows the
                                                                                                                                                                                                                                                                                                                                     selection from among library members of a plurality of that specifically nembers of a plurality of that specifically the molecular complex) 1: the selection from among library members of a plurality of that specifically nonders in the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from among library members of a plurality of the selection from a plurality o
                                                                                                                                                                                                                                                                                                                                                                   bind to target molecule (or molecular complex) is then determined.

then determined binders and to the the chemical building determined.
                                                                                                                                                                                                                                                                                                                                                      molecules (hereinafter called "binders") that specifica molecular complex) 1; the molecular complex) arget molecule structure (e.a., sequence, structure bind to building block structure)
                                                                                                                                                                                                                                                                                                                                                                                      chemical building block structure if predetermined binders and if then are already available. The invention can be their structure are already available.
                                                                                                                                                                                                                                                                                                                                                      formula) is then determined.

The predetermined binders and use the invention can use the invention directly without the need for library the need for library without the need for library without the need for library the need for library without the need for library the need f
                                                                                                                                                                                                                                                                                                                                                                                                                         their structure are already without the need for more their information directly without the done. one or more this information of this screening is done.
                                                                                                                                                                                                                                                                                                                                                                                                                                                    screening is done or more share a library screening is done of binders all share a rne selected binders recific their specific allowing their specific allowing their specific libraries may be screened.

Libraries may be structure. allowing their specific allowing their 
                                                                                                                                                                                                                                                                                                                                                                                                                                      this information directly without the need for library without the need for more this information directly without the selected binders all this screening is done, binders all this screening. The screened
                                                                                                                                                                                                                                                                                                                                                                                                                                                                       libraries may be screened.

The selected binders all sha their specific allowing their scally sing chemically and physically singlibraries may be screened.

The selected binders all sha allowing their specific allowing the
                                                                                                                                                                                                                                                                                                                                                                                                                                      common pharmacophore structure is preferably iteratively structure is preferably iteratively the target in a chemically and physically iteratively in a chemically and physically similar allowing their specific
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       manner.

This common structure is preferably iteration candidate

This common structure is preferably iteration.

The common structure is preferably iteration.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               determined by a selection, geometric and conformational pharmacophore selection, geometric and conformational pharmacophore homologies.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Pharmacophore selection, geometric and conformational structure homologies.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             Geometric and conformational and is

Geometric and conformational and is

this step and is

used at this step and is

and conformational

this step and is

                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               A candidate pharmacophore shared
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                Preferably not considered. A candidate pharmacophore sha for structure is selected. The binders will preferably not by subsequent steps.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                determination by subsequent candidate chemical pharmacophores are typically present ion considerations.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                typically present several candidate chemical pharmacophores are typically present several considerations.

These candidates are these confirmation considerations.

These candidates are these candidates are these confirmations.

These candidates are the confirmation considerations.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 ignoring conformation considerations. These candidates are the signoring conformation building blocks, often contiguous, the signoring conformation building being homologous to the small groups of library building being homologous to the small groups of around in one binder being homologous around in one binder being homologous to the small groups of around in one binder being homologous to the small groups of around in one binder being homologous to the small groups of around in one binder being homologous to the small groups of the small 
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        preferably not considered.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 small groups of library building blocks, homologous to the binder being homologous to the each candidate
```

candidate groups in all the other binders. Building block homologies are determined by applying rules appropriate to the diversity library. In the preferred embodiment, homologous building blocks have similar surface chemical groups, since pharmacophores are defined by a similar geometric arrangement of chemical structures. In the case of the preferred library, CX₆C, candidate pharmacophores are amino acid sequences whose side chain surface groups have similar chemical properties. Amino acid homologies are determined by mechanical rules described below. These candidate sequences are typically 3 amino acids long, but may range from 2 all the way to 6. Where pharmacophores are defined by their charge distributions, homologous library building blocks must have similar charge distributions.

Having selected N binders by screening one or more libraries and determined a candidate pharmacophore in each binder, the subsequent steps of distance measurement, step 4, and Monte Carlo structure determination, step 5, determine a highly accurate structure for the candidate pharmacophore, if 20 possible. This determination will be possible if the candidate is the actual pharmacophore. A subsequent test, step 6, checks for success of this structure determination. In particular cases, distance measurements may not be necessary in order to determine an adequately precise 25 pharmacophore structure.

Measurements are made, step 4, of a few strategic distances in the binders, that will be most useful for the subsequent structure determination step. A minimum number of strategic interatomic distances in the binders are measured 30 in step 4. These few distances constrain possible binder structures and make the subsequent complete structure determination more efficient and more accurate. In preferred but not limiting embodiments, measurement methods yielding distances accurate to at least approximately 0.25 Å or less are used. The preferred methods use nuclear magnetic resonance ["NMR"] techniques. Particularly preferred is the rotational-echo double resonance ["REDOR"] NMR method for

directly measuring ¹³C-¹⁵N internuclear distances in peptides, the most accurate current method for simply and inexpensively obtaining such distances. It is generally capable of accuracy to 0.1 Å and a span of 8 Å. In a specific 5 embodiment, peptide binders are synthesized from amino acids labeled with ¹³C and ¹⁵N. Labeling is chosen to obtain the most useful distance data about the selected candidate pharmacophore structures. Either backbone nuclei, side chair nuclei, or both can be labeled. The step is detailed below.

- 10 Liquid NMR techniques can also be used to indirectly determine internuclear distances in peptides, but are less preferred since they require considerable data interpretation to obtain distances of less accuracy than those obtained by use of REDOR.
- 15 Structure determination, step 5, determines a precise geometric conformation for both the candidate shared chemical structures, if possible, and the remainder of the binders. The preferred but not limiting method, consensus, configurational bias, Monte Carlo ["CCBMC"] determination,
- 20 step 5, is an efficient smart Monte Carlo method uniquely able to incorporate knowledge from prior steps to obtain highly accurate physical binder structures. From library screening, step 2, it is deduced that the binders have a shared, actual pharmacophore, structure because they all bind
- 25 specifically to the same target molecule (hence, a "consensus" method). It is not significant to the method if the binders come from more than one library as long as they all have a structure adaptable to representation in the consensus structure determination step (see infra). From
- 30 distance measurements, step 4, a few strategically chosen distances are accurately known. This information is heuristically utilized along with an accurate model of the physical atomic interactions and the allowed molecular conformations.
- Further, these means are particularly adapted for determining structures of molecules having limited conformational degrees of freedom at the temperature of

interest and conformationally constrained by, e.g., internal bonds. Potential conformations are generated and selected by smart configuration bias techniques which avoid generation of unnecessarily improbable new conformations. (Hence, a 5 "configuration bias" method.) The technique is preferably applied herein to conformationally constrained peptides. A concerted rotation technique is combined with configurational bias conformation generation so that new conformations automatically preserve the internally linked backbone 10 structure constraints. This technique is preferably applied to the preferred constrained peptide library, of a sequence comprising CX₆C (wherein X is any amino acid). The technique is also applicable to other constrained peptide libraries, to peptoid libraries, and to any more general organic diversity 15 libraries that meet certain geometric limitations (i.e., that have structures adaptable to representation in the consensus structure determination step (see infra)).

The methods of the invention are not limited to the use of CCBMC for determining a consensus pharmacophore structure. 20 Alternative embodiments of this invention may use alternative structure determination methods to determine a consensus pharmacophore structure. For example, a simple yet expensive method is to make exhaustive REDOR NMR measurements characterizing the candidate pharmacophore in each binder and 25 then average these measurements. A somewhat less expensive method is to use a conventional Monte Carlo molecular structure determination method to limit somewhat the number of REDOR NMR measurements required to characterize the candidate pharmacophore. Conventional Monte Carlo methods, 30 being unable to directly make use of partial distance measurements or consensus binding information, are less efficient than the CCBMC method and require more distance measurements. Further, other known techniques of molecular structure determination, for example folding rules or 35 molecular dynamics, can be used in place of conventional Monte Carlo.

The success of the structure determination is tested, step 6, against various convergence and success criteria. Consistency tests, step 6, are applied to the resulting structure to determine whether the candidate pharmacophore 5 previously selected is the actual pharmacophore. One set of tests checks predicted distances against new distance measurements or against previous measurements temporarily not used as structure constraints. A second set of tests checks heuristically whether the candidate pharmacophore exhibits 10 the expected low energy consensus structure. The test are described further below. If a shared structure is found, the candidate pharmacophore must be the actual pharmacophore. If not, another candidate pharmacophore and another shared structure is determined, if possible. An actual 15 pharmacophore exists and will eventually be found and accurately structured.

Upon passing these tests, the methods of the invention have provided a consensus structure for the selected candidate pharmacophore, preferably accurate to at least 20 approximately 0.25-0.50 Å, as well as structures for the remainder of the binder molecules. Lead compound selection, step 7, uses these structures to determine or select highly targeted lead compounds 8. One method of lead selection is to design new organic molecules of pharmacologic utility with 25 the determined pharmacophore structure. Another method selects leads from databases of molecular descriptions. Conventionally known to medicinal chemists are databases of potential drug compounds indexed by their significant chemical and geometric structure (e.g., the Standard Drugs 30 File (Derwent Publications Ltd., London, England), the Bielstein database (Bielstein Information, Frankfurt, Germany or Chicago), and the Chemical Registry database (CAS, Columbus, Ohio)). The determined pharmacophore, being a chemical and geometric structure in the preferred embodiment, 35 is used to query such a database. Search results will be those compounds with homologous chemical groups arrayed in a very closely similar geometric arrangement. These are lead

compounds 8 output from this invention and input to the process of drug testing and development.

Although the preferred identity and ordering of the method steps is presented in Fig. 1, the invention is not 5 limited to this identity and ordering. Other orderings, especially of steps 3, 4, and 5, are possible to achieve certain efficiencies. Steps can be inserted and deleted, for optimal effect. For example, an additional partial structure determination step can be inserted between existing steps 3 10 and 4 to provide information on how best to make the step 4 strategic measurements. As another example, in an alternative aspect, in lieu of screening one or more libraries to select binders, predetermined binders can be obtained and used (e.g., binders determined by any means to 15 be specific to the same target molecule); thus, step 2 can be omitted. In another embodiment, step 4, the measurement step, can be omitted. While all method steps in the preferred embodiment assume an aqueous environment at body temperature (37 °C), to the extent these parameters are 20 relevant to the particular step, the invention is not limited to human environmental parameters.

Screening against a diversity library consists of selecting by assay those library members which bind specifically to the target molecule of interest. Binding 25 specificity is preferably a binding constant of less than 1 µm (micromolar), and more preferably less than 100 nm (nanomolar). Preferably, an assay is done that detects an effect of binding of the binder to the target molecule on the target molecule's biological activity, to ensure that the 30 binding is actually to the biological target of interest. Also, preferably, the selected binders are tested to further select those binders that bind to the target molecule competitively, to ensure that each binds to the same target in the target molecule.

The output of the screening step is a number, N, of binders selected from one or more libraries for use by the subsequent steps of the method. The binders with highest

```
affinity are preferably selected for use by the N himhers affinity the chemical structure of each of the chemical structure of each of the chemical structure of the chemical 
                                                                                                                               steps. The chemical structure of each of the member the nemical structure of each of the member of each of the member of each of the member of the nemical structure of each of the member of each of the member of the nemical structure of each of the member of each of the member of the mem
                                                                                                                                            selected for use is determined as part of the member that the primary chemical intrary screening.

Selected for use is determined onstrained peoride library screening.

Synthesis and library preferred constrained peoride library screening.
                                                                                                                               synthesis and library screening.

The primary chemical is the preferred constrained peptide and nortion of the preferred constrained of the amino acid secuence of the structure of the amino acid secuence of the acid sec
                                                                                                                                                                           structure of the preferred constrained peptide library is of the preferred constrained of the Aiversity specified by the amino more deneral ordanic diversity specified by molecule.
                                                                                                                                                                                                       the CX+C molecule. For more general organic diversity building arrangement of library building the Selection and arrangement.

Libraries, the binders must be determined.
WO 96130849
                                                                                                                                                                                                                                                                                                               ks in the binders must be determined. that the set of this invention and small the frament of the serious and small the set of this invention that frament is selective and small the set of this invention that frament is selective and small the set of this invention that frament is selective and small the set of this invention that frament is selective and small the set of this invention that frament is selective and small the set of this invention that frament is selective and small the set of this invention that frament is selected as preferred as preferred
                                                                                                                                                                                                                                                                      cf determined lead compounds is selective and small. Example tolerances are distance tolerances are distance tolerances are tolerances are tolerances as pharmacophore distance by Armo Aarahaee that as pharmacophore distance illustrates number of compounds retrieved by Armo Aarahaee that as pharmacophore distance that as pharmacophore distance to the number of compounds retrieved by Armo Aarahaee that as pharmacophore distance to the pharmacophore dista
                                                                                                                                                                                                                         IlDrarles, the binders must be determined.

blocks in the binders must be determined.
                                                                                                                                                                                                                                                      It is a preferred aspect of this invention that invention the invention that invention the invention that invention that invention the invention that invention the invention that inventi
                                                                                                                                                                                                                                                                                  lillustrates that as pharmacophore distance tolerances are tolerances darabase that as pharmacophore distance by drug darabase as pharmacophore distance by drug darabase invention as this invention as this invention as the number of compounds as this invention as the number of compounds as the number of compounds.
                                                                                                                                                                                                  the CX,C molecule.
                                                                                                                                                                                                                                                                   searches increases geometrically.

Searches ingh resolution searches, or other methods of that database searches be expected that
                                                                                                                                                                                                                                                                                                                            determines high resolution pharmacophore geometries, it con the pharmacophore geometries, of the database searches, or other methods of will resolution pharmacophore structure, will resolute that database searches, or other methods of the pharmacophore structure, will resolute the expected that database pharmacophore structure, will resolve the expected that database pharmacophore structure, will resolve the expected that database pharmacophore structure.
                                                                                                                                                                                                                                                                                                                                              be expected that database searches leads.

be expected that leads from pharmacophore targeted leads.

determining selective.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Methods limiting the
                                                                                                                                                                                                                                                                                                                                                                        only a few, selective, the cost of the utility to the pharmaceutical number of considerable utility to consequently of consequ
                                                                                                                                                                                                                                                                                                                                              number of leads decrease the cost of drug development and ar to the pharmaceutical

number of leads considerable utility to the expense of developing the considerable utility. The expense of developing the consequently of considerable community.

20 consequently and medical community industry and medical community.
                                                                                                                                                                                                                                                                                                                                                                                                                            industry and medical community for biological effect and romnound and evaluating lead compounds known.
                                                                                                                                                                                                                                                                                                                                                                                                                                      and evaluating lead compounds known.

and evaluating lead compounds known.

medicinal usefulness is well usefulness. efficacv.

medicinal usefulness pharmacological usefulness.
                                                                                                                                                                                                                                                                                                                                                                                                                                                        medicinal usefulness is well known.

pharmacological usefulness, reduired and the medicinal usefulness are reduired and the medicinal usefulness.

safetv.
                                                                                                                                                                                                                                                                                                                                                                                                                            be screened for pharmacological modifications are required in vivo finally, the required in pharmacological pharmacological pharmacological pharmacological with the required in vivo finally, the required in pharmacological modifications are required in vivo pharmacological pharmacologi
                                                                                                                                                                                                                                                                                                                                                                                                                  industry and medical community.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     often chemical modifications are required in vivo

Finally, trials alone can cons

process must toxicity and clinical trials

process process pharmacologic
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  Process must be repeated and clinical trials alone can consume pharmacologic and millions of dollars.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  of time and millions of dollars molecule 1 having a target molecule target the time and millions with a target interesting target the three fore, pharmacologically interesting target three for the target three dollars interesting target three dollars.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          Therefore, starting with a target molecule 1 having a conservation of this invention determines a conservation biologically or pharmacologically of this invention determines a conservation determines and determines a conservation determines a conservat
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                biologically or pharmacologically interesting target, the interesting target, consensus biologically or pharmacologically interesting target, inte
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                method and apparatus of this invention determines a consensus pharmacophore rhat consensus pharmacophore as selective new pharmacophore then be used to determine a selective pharmacophore then be used to determine a selective new pharmacophore then be used to determine a selective new pharmacophore then be used to determine a selective new pharmacophore then be used to determine a selective new pharmacophore then be used to determine a selective new pharmacophore and apparatus of this invention determines a consensus pharmacophore new pharmacopho
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  Pharmacophore structure.

This consensus pharmacophore get of determine a selective get determine a for rational deciments and to determine a for rational deciments and then be used to determine a for rational deciments and then be used commonwhere a formation and the structure can be an amount of the highly specific lead commonwhere a formation and the highly specific lead commonwhere a formation and the highly specific lead commonwhere a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific lead to determine a formation and the highly specific le
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            structure can then be used to determine as ligand-mimics (adonists highly specific capable of drugs, sof drugs,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  highly specific lead compounds a (Fig. 1) for rational design as ligand-mimics (agonists nolecule.

highly specific capable of acting as target molecule.

35 of drugs of antagonists for the particular target and acting as target molecule.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              of drugs, e.g., for the particular target molecule.

or antagonists)
```

In the following discussion and examples, each of these steps will be more fully described.

5.1. SELECTION OF A TARGET MOLECULE

- The target molecule is any one or more molecules containing a target or putative target of interest. The target is a binding interaction region. The target can be in a single molecule or can be a product of a molecular complex. The target can be a continuous or discontinuous binding
- 10 region. The target molecule selected for use (Fig. 1, step 1) is preferably any molecule that is found in vivo (preferably in mammals, most preferably in humans) and that has biological activity, preferably involved or putatively involved in the onset, progression, or manifestation of a
- 15 disease or disorder. The target molecule can also be a fragment or derivative of such an in vivo molecule, or a chemical entity that contains the same target as the in vivo molecule. Examples of such molecules are well known in the art. Such molecules can be of mammalian, human, viral,
- 20 bacterial, or fungal origin, or from a pathogen, to give just some examples. The target molecule is preferably a protein or protein complex. The target molecules that can be used include but are not limited to receptors, ligands for receptors, antibodies or portions thereof (e.g., Fab, Fab',
- 25 F(ab')₂, constant region), proteins or fragments thereof,
 nucleic acids, glycoproteins, polysaccharides, antigens,
 epitopes, cells and cellular components, subcellular
 particles, carbohydrates, enzymes, enzyme substrates,
 oncogenes (e.g., cellular, viral; oncogenes such as ras, raf,
- 30 etc.), growth factors (e.g., epidermal growth factor, platelet-derived growth factor, fibroblast growth factor), lectins, protein A, protein G, organic compounds, organometallic compounds, viruses, prions, viroids, lipids, fatty acids, lipopolysaccharides, peptides, cellular
- 35 metabolites, steroids, vitamins, amino acids, sugars, lipoproteins, cytokines, lymphokines, hormones, T cell surface antigens (e.g., CD4, CD8, T cell antigen receptor),

ions, organic chemical groups, viral antigens (hepatitis B
virus surface or core antigens, HIV antigens (e.g., gp120,
 gp46)), hepatitis C virus antigens, toxins (e.g., bacterial
 toxins), cell wall components, platelet antigens (e.g.,
5 gpiibiiia), cell surface proteins, cell adhesion molecules,
 neurotrophic factors, and neurotrophic factor receptors.

In specific embodiments, vEGF (vascular endothelial growth factor) or KDR (the receptor for vEGF) (Terman et al., 1992, Biochem. Biophys. Res. Comm. 187:1579-1586) is the 10 target molecule. vEGF and its receptor are the major regulators of vasculogenesis and angiogenesis (Millauer et al., 1993, Cell 72:835). Inhibition of the vEGF and the concomitant inhibition of its mitogenic activity and angiogenic capacity has been shown to suppress tumor growth 15 in vivo (Kendall et al., 1993, Proc. Natl. Acad. Sci. USA 90:10705-10709; Kim et al., 1993, Nature 362:841-844). Use of vEGF or KDR or portions thereof, as a target molecule is a preferred embodiment for use of the present invention to develop lead molecules as drugs in the area of cardiovascular 20 disease or cancer.

The proteins ras and raf, or portions thereof (e.g., modules -- functional portions), are also preferred target molecules, particularly in an embodiment wherein the methods of the present invention are employed to develop lead 25 molecules for drugs that are cancer therapeutics. ras is a member of an intracellular signaling cascade that controls cell growth and differentiation (Cook and McCormick, 1994, Nature 369:361-362). ras functions in signal transduction by specifically recognizing the protein raf and bringing it to 30 the cell membrane (Hall, 1994, Science 264:1413-1414; Vojtek et al., 1993, Cell 74:205-214). The recognition modules in both ras and raf have been determined (Zhang et al., 1993, Nature 364:308-313; Warne et al., 1993, Nature 364:352-355; and Vojtek et al., 1993, Cell 74:205-214); in a specific 35 embodiment, such a recognition module is used as a target molecule according to the invention.

```
In another specific embodiment, an integrin is used as a integrin is used as a integrin is used as a function in
                                                         target molecule. Such molecules are known to the present of the present and can be used according to the present of the present of the present to the presen
                                                               clot formation, and can be used according to the present the area of invention to develop lead molecules for drugs in the airea of invention to develop lead molecules.
                                                                                                                      lovascular disorders.

Target molecules commercially available or can be obtained commercially available or can be obtained commercially available.
                                                                                           Target molecules for use can be obtained commercially available, or can be commercially available, or recombinant sour
                                                                                               synthesized or purified from natural or recombinant sources.

The specific embodiment, a specific embodiment, and a specific embodiment, and a specific embodiment, and a specific embodiment.
WO 96130849
                                                      target molecule.
                                                                                       In a specific embodiment, a target molecule is prepared that a target molecule is prepared that a rather an "affinity binding partner in a specific embodiment, a target molecule is prepared that incorporate an "affinity binding partner in a specific embodiment, a target molecule is prepared that in a target molecule is prepared to a "affinity binding partner in a target molecule is prepared to a "affinity binding partner in a target molecule is prepared to a known binding partner in a specific embodiment, a target molecule is prepared to a known binding partner in a specific embodiment, a target molecule is prepared to a known binding partner in a target molecule is prepared to a known binding partner in a target molecule is prepared to a target molecule is prepare
                                                                                                                       has been modified to incorporate an "affinity tag," i.e., a target to incorporate an "affinity binding partner, binds to a known binding of the target to facilitate recovery/isolation/immobilization of the target structure that recovery/isolation/immobilization of the target structure recovery/isolation/immobilization of the target recovery/isolation/immobilization of the target recovery/isolation/immobilization of the target recovery/isolation of the tar
                                                             s cardiovascular disorders.
                                                                                                                              structure that specifically binds to a known binding partner that specifically binds to a known binding partner to a known binding partner to facilitate recovery isolation immobilization of the target exoression recombinant exoression recombinant exoression recombinant exoression aspect.
                                                                                                                                      to facilitate recovery isolation in the art can be used to produce a molecule.

molecule well known in the art can be used to molecule.
                                                                                                                              methods well known in the as a ffinity tags include but are not such affinity tags.

15 protein target molecule as a ffinity tags.
                                                                                                                                                            protein target molecule such antibodies (e.a., c-mvc enirone affinity tags (e.a., c-mvc enirone affinity tags)

protein affinity tags of known antibodies (e.a., c-mvc enirone affinity tags)
                                                                                                                                                                  peptide affinity tag. Known antibodies (e.g., c-myc epitopes of known Biol. s.3610-3616)), a series limited to epitopes Mol. (Evan et al.,
                                                                                                                                                                             limited to epitopes of known antibodies (e.g., aseries was to epitopes of known antibodies (e.g., altose limited to epitopes Mol. Cell. (which bind to zinc). maltose (evan et al., of his residues (which bind to zinc).
                                                                                                                                                                                      (Evan et al. of his such as pmal. etc. 1985 such as pmal. etc. (e.g., binding sequences such as pmal.)
                                                                                                                                                                     (e.g., sequences at either the amino or carboxy. terminus into protein targets at either the amino or carboxy.
                                                                                                                                                                                                   binding sequences such as pmal, etc. Tags are incorporated to a either the amino or carboxy-terminus.

binding sequences at either is chemically attached to a the tardet is chemically attached to a the tardet in another embodiment.
                                                                                                                                                                                                         into protein targets at either the amino or carboxy-terminus.

In another embodiment, binds to avidin.

tag (e.g., biotin (which binds to avidin))
                                                                                                                                                                                                                   In another embodiment, the target avidin, streptavidin) the target avidin, streptavidin, binds to avidin, streptavidin, by biotinvlation tag (e.g., biotin by biotinvlation) tag (e.g., biotin by biotinvlation)
                                                                                                                                                                                                                                                                         ptavidini e.g., by blotlinylation. by standard methods.

The target a protein rarger can be mirified by crandard for the protein rarger can be mirified by crandard for the target appropriate target appro
                                                                                                                                                                                                                                           The target molecule is purified by standard methods.

The target molecule is purified by standard methods including chromatography (e.g., ion exchange)

For example, a protein chromatography methods including chromatography
                                                                                                                                                                                                                                                        methods including chromatography or by any other standard technique affinity affinity differential
                                                                                                                                                                                                                              For example, and sizing column chromatography and sizing column chromatography
                                                                                                                                                                                                                                                                  affinity and sizing column chromatography) centrifugation, chromatography) centrifugation of by any other standard embodiment.

affinity and solubility or by any other preferred embodiment.

differential solubility of proteins: in a preferred embodiment.
                                                                                                                                                                                                                                                  differential solubility or by any other standard technique or by any other preferred embodiment, or by any other preferred embodiment, in a preferred chromatography)

differential solubility proteins; in a liquid chromatography

differential solubility of proteins; in a liquid chromatography

differential solubility proteins; in a liquid chromatography

the purification (high performance liquid chromatography)
                                                                                                                                                                                                                                                                                 for the purification of proteins; in a preferred embodiment, in the purification of proteins; in a preferred embodiment, in the purification of proteins; in a preferred embodiment, in the purification of proteins; in a preferred embodiment, in the purification of proteins; in a preferred embodiment, in the purification of proteins; in a preferred embodiment, in the purification of proteins; in a preferred embodiment, in the purification of proteins; in a preferred embodiment, in the purification of proteins; in a preferred embodiment, in the purification of proteins; in a preferred embodiment, in the purification of proteins; in the purification of proteins and proteins are proteins.
                                                                                                                                                                                                                                                                                                                                       nployed.

Once the target molecule has been purified, it is
                                                                                                                                                                                                                                                                                                         once the target molecule has been purified, it is biological and once the target ensure that it retains its native conformation.

preferably tested to retains its native conformation.
                                                                                                                                                                                                                                                                                                                        activity (and thus or in vivo assay is a fracment or derivative suitable in view the desired target molecule is where the desired
                                                                                                                                                                                                                                                                                         preferably tested to ensure that it retains its biolog.

In the preferably tested to ensure that it retains its native conformation).

Preferably tested to ensure that it retains its native conformation).

In the preferably tested to ensure that it retains its proposed in the preferable that it retains it retains it retains its native conformation).

Preferably tested to ensure that it retains its biolog.

In the preferably tested to ensure that it retains its biolog.

In the preferably tested to ensure that it retains its biolog.

In the preferably tested to ensure that it retains it re
                                                                                                                                                                                                                                                                                                                               Euitable in vitro or in vivo assay can be used. In instances or derivative is a fragment or derivative where the desired target
```

of a molecule found in vivo, or is a chemical entity putatively containing the same target as a molecule found in vivo, it is highly preferred that testing be done of such desired target molecules prior to their use, so that among 5 such desired target molecules, only those that have the same biological activity as the in vivo molecule or compete with a known ligand to the in vivo molecule, are selected for actual use as target molecules according to the invention. In the event that biological activity has been reduced or lost in a 10 recombinant protein relative to the native form of the protein, the protein can be recombinantly expressed in a different host (e.g., yeast, mammalian, or insect) and/or with a variety of tags and location of tags (on either the amino- or carboxy-terminal side), in order to attempt to 15 achieve, or to optimize, recovery of biological activity.

5.2. DIVERSITY LIBRARIES

According to a preferred embodiment of the invention, diversity libraries are screened to select binders, which 20 specifically bind to the target molecule. Diversity libraries are those containing a plurality of different members. Generally, the greater the number of library members and the greater the probability that all possible members are represented, the more preferred the library. In preferred embodiments, the diversity libraries have at least 10⁴ members, and more preferably at least 10⁶, 10⁸, 10¹⁰, or 10¹⁴, members.

Many libraries suitable for use are known in the art and can be used. Alternatively, libraries can be constructed

30 using standard methods. Chemical (synthetic) libraries, recombinant expression libraries, or polysome-based libraries are exemplary types of libraries that can be used.

In a preferred embodiment, the library screened is a constrained, or semirigid library (having some degree of 35 structural rigidity). Examples of constrained libraries are described below. A linear, or nonconstrained library, is

```
Additionally, one or
                                                                                less preferred although it may be used. Additionally, one on a select binders.

The although it may be used. Additionally, one on a select binders. The library contains mention the library m
                                                                                                                                                                              different libraries can be screened to select binders. Amin's the library contains peptide the library contains peptide amin's the library contains peptide embodiment. In the range of select binders amin's the range of select binders. The figure of select binders amin's contains peptide embodiment. In the range of select binders. The range of select binders amin's contains peptide embodiment.
                                                                                                              In a preferred embodiment, the library member.

In a preferred having a length library member.

or peptide analogs thereof in each library member.
                                                                     less Preferred although it may be used.
                                                                                                                                                                                                                            thereof in each library member ified from a jdentified from a binders are identified from a bind
                                                                                                                                                           In specific embodiments, binders are identified from a library or a chemically synthesized libraries term "random" peptide expression the term random peptide library.
                                                                                                   or peptide analogs thereof in each library member.

S acids or analogs thereof analogs there are analogs thereof analogs there
                                                                                                                                                                           random peptide expression library or a chemically synthesized libraries

The term "random" peptide of both its scope libraries of both random peptide library within its scope library to include within its scope libraries of both random peptide library to include within its scope li
WO 96130849
                                                                                                                                                                                              random peptide library.

The term "random" peptide library of both

The term "random" peptide both

its scope library.

(variant) nentides

is meant to include within its (variant) nentides

is meant to and rotally random

is meant and rotally random
                                                                                                                                                                                                                                                                                                          ially and totally random the peptide rhar are chamically to libraries that are chamically in the peptide rhar are chamically to peptide rhar are chamically and totally the peptides rhar are chamically to peptides.
                                                                                                                                                                 is meant to include within its scope libraries of (variant) peptides.

Is meant to include within its scope libraries is to include within its scope libraries of variant) peptides.

To partially and totally random rentide libraries is to partially and rentiment research to partially and rentiment rentide libraries is to partially and rentiment rentide libraries of the peptides.
                                                                                                                                                                                                                                            In one embodiment, the peptide libraries used in the peptide that are chemically the peptide t
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       De Libraries of such libraries are given
                                                                                                                                                                                                                                                                            synthesized in vitro. Examples of such libraries are given an synthesized in vitro. Science 251:167-173, which describes on an synthesized al., a known array of short neutides on an in Fodor et al., a known array of the synthesis of of the synthes
                                                                                                                                                                                                                                            in Fodor et al. 1991 Na array of short pertides 1991 Na in Fodor et al. 1991 Na array houghten et al. 1991 Na 15 the synthesis microscopic slide:
                                                                                                                                                                                                                                                                                                            the synthesis of a known slide; hort pertides on an individual which describes mixtures of tree hexapentides in individual which describes mixtures of the synthesis of describes mixtures of the hexapentides in individual which describes mixtures of the hexapentides in t
                                                                                                                                                                                                                                                                                                                          individual microscopic slide; Houghten et al.; Houghten e
                                                                                                                                                                                                                                                                                                                                          354:84.86, which describes mixtures of free hexapertides in each peptide were and second residues in each er all which the first and specifically defined:

which the first and specifically defined:

individually and specifically defined:
                                                                                                                                                                                                                                                                                                                                                       which the first and specifically defined; none head one nent; individually and specifically describes a none head one he
                                                                                                                                                                                                                                                                                                                       individually and specifically defined; Lam et al., peptide"

individually and specifically describes a "one bead, one peptide"

individually and specifically describes a synthesis scheme

individually and specifically defined; Lam et al., peptide"

individually and specifically describes a "one peptide"

individually and specifically approach in which a solid phase split specifically describes a "one peptide"

individually and specifically approach in which a solid phase split specifically approach in which a solid phase split specifically approach in which a specific specifically approach in which a specific specifically approach in which a specific specific
                                                                                                                                                                                                                                                                                                                                                                                          Nature 354:82-84, which a solid phase in which each head in the approach in which a library of people approach a library of people approach a library control approach a library contro
                                                                                                                                                                                                                                                                                                                                                                                                     approach in which a solid phase in which each bead in semi-
                                                                                                                                                                                                                                                                                                                                                                                                                     Produced a library of peptides in which each bead in the single, random sequence thereon a single, Bio/Technology ecolection had immobilized Medynski, 1994. Bio/Technology of amino acid residues:
                                                                                                                                                                                                                                                                                                                                                                                                                                     collection had immobilized thereon 1994, Bio/Technology

which describes solit synthesis and T-bad

of amino acid which describes solit

i. 72.704-710
                                                                                                                                                                                                                                                                                                                                                                                                        of amino acid residues: Medyneki, 1994, Bio/Technology

of amino acid residues: Medyneki, 1994, J. Medicir

and Gallop et al., 1994

25.12:709-710, mernods:
                                                                                                                                                                                                                                                                                                                                                                                                                                                                        12:709-710, which describes split synthesis and T-bag which describes split synthesis and of other examn simply by way of other examn synthesis methods; and Gallop simply by way of other examn chemistry 37(9):1233-1251.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      synthesis methods: and Gallop et al., le prepared for use. according to themistry and library may be prepared for use. according to themistry and library may be prepared for use.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   Chemistry 37(9):1233-1251. Simply by way of other examples to simply by way of other examples to be prepared for use, Acad. Sci.

Chemistry 37(9):1233-1251. De prepared proc. Natl. Acad. Sci.

a combinatorial ohlmeyer et al., 1993, proc. hatl. the methods of ohlmeyer et al., 1993, proc.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      a combinatorial library may be prepared for wath. Acad. Sci.

the methods of Onlmeyer et al., 1994, proc. wath. Acad. Sci.

1984, proc. Math. Acad. Sci.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      the methods of Onlmeyer et al., 1994, 1992, Biotechniques
30 USA 91:11422-11426; Houghten et al., 1994, 91:11422-11426; Houghten et al., 1992, Biotechniques
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    USA 90:10922-10926; Erb er al., 1994, Proc. Natl. Acad. Sci. Houghten 1994, Proc. Natl. Acad. Sci. Houghten 1994, Proc. Natl. Acad. Sci. Houghten 1994, Proc. Natl. Acad. Sci. 13:412; Javawickreme et al., 1994, Proc. Natl. Acad. Sci. 13:412; Javawickreme et al., 1994, Proc. Natl. Acad. Sci. Natl. Natl. Acad. Sci. Natl. Natl. Acad. Sci. Natl. Aca
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    USA 91:11422-11426; Houghten et al., 1993, Proc. Natl. Acad. sci.

13:412; Jayawickreme et al., 1993, Proc. Natl. Acad. 91:1614-1618; Or Salmon et al., 1993, Proc. Natl. Acad. 91:1614-1618;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    13:412; Jayawickreme et al., 1994, Proc. Natl. Acad. Sci. Watl. Acad. Natl. Acad. Natl. Acad. Natl. Acad. Natl. Acad. Natl. Acad. Natl. Acad. No. Natl. Acad. No. No. Natl. Acad. Natl. Acad. Natl. Acad. Natl. Acad. Natl. Natl. Natl. Acad. Natl. Natl
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    USA 90:11708-11712. 1992. Per Publication No. WO 93/20242 a proc. Natl. Acad. chemical proc. Natl. Acad. chemical proc. Salar and Lerner. 1992. Mancoded combinatorial chemical salar and Lerner. Per Mancoded combinatorial chemical processing and percentage of the mancoded combinatorial processing and percentage of
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Brenner and Lerner, 1992, wencoded combinatorial chemical wencoded combinatorial chemical 89:5381-5383 describe
```

libraries," that contain oligonucleotide identifiers for each chemical polymer library member.

In another embodiment, biological random peptide libraries are used to identify a binder which binds to a 5 target molecule of choice. Many suitable biological random peptide libraries are known in the art and can be used or can be constructed and used to screen for a binder that binds to a target molecule, according to standard methods commonly known in the art.

According to this approach, involving recombinant DNA techniques, peptides are expressed in biological systems as either soluble fusion proteins or viral capsid fusion proteins.

In a specific embodiment, a phage display library, in 15 which the protein of interest is expressed as a fusion protein on the surface of a bacteriophage, is used (see, e.g., Smith, 1985, Science 228:1315-1317). A number of peptide libraries according to this approach have used the M13 phage. Although the N-terminus of the viral capsid 20 protein, protein III (PIII), has been shown to be necessary for viral infection, the extreme N-terminus of the mature protein does tolerate alterations such as insertions. The protein PVIII is a major M13 viral capsid protein, which can also serve as a site for expressing peptides on the surface 25 of M13 viral particles, in the construction of phage display libraries. Other phage such as lambda have been shown also to be able to display peptides or proteins on their surface and allow selection; these vectors may also be suitable for use in production of libraries (Sternberg and Hoess, 1995, 30 Proc. Natl. Acad. Sci. USA 92:1609-1613).

Various random peptide libraries, in which the diverse peptides are expressed as phage fusion proteins, are known in the art and can be used. Examples of such libraries are described below.

35 Scott and Smith, 1990, Science 249:386-390 describe construction and expression of a library of hexapeptides on the surface of M13. The library was made by inserting a 33

PCT/US96/04229

base pair Bgl I digested oligonucleotide sequence into an Sfi I digested phage fd-tet, i.e., fUSE5 RF. The 33 base pair fragment contains a random or "degenerate" coding sequence (NNK), where N represents G, A, T or C and K represents G or 5 T. Cwirla et al., 1990, Proc. Natl. Acad. Sci. USA 87: 6378-6382 also described a library of hexapeptides expressed as pIII gene fusions of M13 fd phage. PCT publication WO 91/19818 dated December 26, 1991 by Dower and Cwirla describes a library of pentameric to octameric random amino 10 acid sequences.

Devlin et al., 1990, Science, 249:404-406, describes a peptide library of about 15 residues generated using an (NNS) coding scheme for oligonucleotide synthesis in which S is G or C.

- 15 Christian and colleagues have described a phage display library, expressing decapeptides (Christian, R.B., et al., 1992, J. Mol. Biol. 227:711-718): The DNA of the library was constructed by use of an oligonucleotide comprising the degenerate codons [NN(G/T)]₁₀ (SEQ ID NO:8) with a self-
- 20 complementary 3' terminus. This sequence forms a hairpin which creates a self-priming replication site that was used by T4 DNA polymerase to generate the complementary strand. The double-stranded DNA was cleaved at the SfiI sites at the 5' terminus and hairpin for cloning into the fUSE5 vector 25 described by Scott and Smith, supra.

Lenstra, 1992, J. Immunol. Meth. 152:149-157 describes a library that was constructed by annealing oligonucleotides of about 17 or 23 degenerate bases with an 8 nucleotide long palindromic sequence at their 3' ends. This resulted in the

- 30 expression of random hexa- or octa-peptides as fusion proteins with the β -galactosidase protein in a bacterial expression vector. The DNA was then converted into a double-stranded form with Klenow DNA polymerase, blunt-end ligated into a vector, and then released as Hind III fragments.
- 35 These fragments were then cloned into an expression vector at the sequence encoding the C-terminus of a truncated β -galactosidase to generate 10 7 recombinants.

Kay et al., 1993, Gene 128:59-65 describes a random 38 amino acid peptide phage display library.

PCT Publication No. WO 94/18318 dated August 18, 1994 describes random peptide phage display "TSAR libraries" that 5 can be used.

Other biological peptide libraries which can be used include those described in U.S. Patent No. 5,270,170 dated December 14, 1993 and PCT Publication No. WO 91/19818 dated December 26, 1991.

- In a specific embodiment, a "peptide-on-plasmid" library, containing random peptides fused to a DNA binding protein that links the peptides to the plasmids encoding them, can be used (Cull et al., 1992, Proc. Natl. Acad. Sci. USA 89:1865-1869).
- 15 Another alternative to phage display or chemically synthesized libraries is a polysome-based library, which is based on the direct in vitro expression of the peptides of interest by an in vitro translation system (in some instances, coupled to an in vitro transcription system).
- 20 These methods rely on polysomes to translate the genomic information (in this case encoded by an mRNA molecule, in some instances made in vitro by transcription from synthetic DNA) (see, e.g., Korman et al., 1982, Proc. Natl. Acad. Sci. USA 79:1844-1848). Such in vitro translation-based libraries
- 25 include but are not limited to those described in PCT Publication No. WO 91/05058 dated April 18, 1991; and Mattheakis et al., 1994, Proc. Natl. Acad. Sci. USA 91:9022-9026.

Diversity library screening, step 2 of Fig. 1,

- 30 determines a few, N, members (compounds) from one or more libraries and their primary sequences all of which specifically bind to target molecule 1 in a similar manner. A structured organic diversity library is a prescription for the creation of a huge number of related molecules all built
- 35 from combinations of a small number of chemical building blocks. Preferred diversity libraries for use according to the invention have members whose binding to a target molecule

is characterized by configurational entropy change that are relatively small to the binding energy. This means that library members have definite structures in the bound and, especially, the unbound states. A preferred example of a 5 chemical diversity library for use in the invention contains short peptides with a constrained conformation. Short peptides without constrained conformations are often freely flexible in an aqueous environment and adopt no fixed unbound structure. The binding of such library members is 10 complicated by significant configurational entropy changes. To eliminate this complication, it is preferred that all library members have a constrained structure and bind to the target molecule in a specific and identifiable manner. One method of achieving constrained conformation is to require

15 internal linking, such as by disulfide bonds. In one embodiment, disulfide bond formation is achieved by use of libraries that contain peptides having a pair of invariant cysteine residues, preferably positioned in the range of 2-16 residues apart, most preferably 6-8 residues 20 apart, that cross-link in an oxidizing environment to form cystines (disulfide bonds between cysteines). An example of such libraries are those containing or expressing peptides of the form R1CX_CR2 wherein R1 is a sequence of 0-10 amino acids, C is cysteine, Xn is a sequence of n variant amino acids 25 (e.g., if all 20 classical amino acids are represented, X means any one of the 20 classical amino acids); n is an integer ranging from 2 to 16; and R2 is a sequence of 0-10 amino acids. R1 and R2 can contain invariant or variant amino acids. Another example is such libraries are those 30 containing or expressing peptides of the form R¹CX_nR², where R^1 , X, n, and R^2 are as described above; n is preferably 8 or 9. A preferred constrained peptide library, of at least 106

A preferred constrained peptide library, of at least 10⁶ members, consists of peptides comprising the sequence CX₆C (SEQ ID NO:1), wherein C is cysteine, X is any naturally
 occurring amino acid, and a disulfide bond is formed between the two cysteines. Additional invariant amino acids (e.g., preferably no more than 5-10 amino acids) on either the

amino- or carboxy-terminus of CX,C can be incorporated as part of the peptide in this preferred embodiment. Fig. 10 schematically illustrates such a molecule. The disulfide bridge between the two cysteines acts as a sufficient 5 conformational constraint for the preferred practice of this invention. By way of example, the library is constructed by generating oligonucleotides with the desired degeneracy to code for the peptides and ligating them into vectors of These inserted oligonucleotides are suitable for 10 both use in in vivo genetic expression systems exemplified by phage display, or in vitro translation methods based on coupled transcription and translation from DNA of interest (see below). The creation and use of an exemplary library is described in Section 6.3 hereinbelow. The invention is 15 easily and readily adaptable to other alternative peptide libraries which include short peptides with alternative disulfide scaffolding, for example, comprising the sequence CX_CX_CC with two disulfide bridges, wherein n and m are each independently an integer in the range of 2-10, and X is any 20 amino acid. More generally, any peptide library containing members of definite conformation which bind to a target molecule in a specific and identifiable manner may be used. Further, more general, structurally constrained, organic diversity (e.g., nonpeptide) libraries, can also be used. By 25 way of example, a benzodiazepine library (see e.g., Bunin et al., 1994, Proc. Natl. Acad. Sci. USA 91:4708-4712) may be

Constrained libraries that can be used are also known in the art. For example, PCT Publication No. WO 94/18318 dated 30 August 18, 1994 describes semirigid phage display libraries, in which the plurality of expressed peptides can adopt only a single or a small number of conformations. Examples of such libraries have a pair of invariant cysteine residues positioned in or flanking random residues which, when 35 expressed in an oxidizing environment, are most likely crosslinked by disulfide binds to form cystines. Also disclosed are libraries having a cloverleaf structure by appropriate

adapted for use.

arrangement of cysteine residues. Also disclosed are libraries with peptides having invariant cysteine and histidine residues positioned within the random residues, or invariant histidines alone within the random residues.

5 TSAR-13 and TSAR-14 are exemplary semirigid libraries disclosed therein.

Other conformationally constrained libraries that can be used include but are not limited to those containing modified peptides (e.g., incorporating fluorine, metals, isotopic

10 labels, are phosphorylated, etc.), peptides containing one or more non-naturally occurring amino acids, non-peptide structures, and peptides containing a significant fraction of γ-carboxyglutamic acid.

As stated above, libraries of non-peptides, e.g.,

- 15 peptide derivatives (for example, that contain one or more non-naturally occurring amino acids) can also be used. One example of these are peptoid libraries (Simon et al., 1992, Proc. Natl. Acad. Sci. USA 89:9367-9371). Peptoids are polymers of non-natural amino acids that have naturally
- 20 occurring side chains attached not to the alpha carbon but to the backbone amino nitrogen. Since peptoids are not easily degraded by human digestive enzymes, they are advantageously more easily adaptable to drug use. Another example of a library that can be used, in which the amide functionalities
- 25 in peptides have been permethylated to generate a chemically transformed combinatorial library, is described by Ostresh et al., 1994, Proc. Natl. Acad. Sci. USA 91:11138-11142).

The peptide or peptide portions of members of the libraries that can be screened according to the invention are not limited to containing the 20 naturally occurring amino

- 30 not limited to containing the 20 naturally occurring amino acids. In particular, chemically synthesized libraries and polysome based libraries allow the use of amino acids in addition to the 20 naturally occurring amino acids (by their inclusion in the precursor pool of amino acids used in
- 35 library production). In specific embodiments, the library members contain one or more non-natural or non-classical amino acids or cyclic peptides. Non-classical amino acids

include but are not limited to the D-isomers of the common amino acids, α-amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid; γ-Abu, ε-Ahx, 6-amino hexanoic acid; Aib, 2-amino isobutyric acid; 3-amino propionic acid; 5 ornithine; norleucine; norvaline, hydroxyproline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β-alanine, designer amino acids such as β-methyl amino acids, Cα-methyl amino acids, Nα-methyl amino acids, fluoro-amino acids and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

By way of example, the incorporation of non-standard or modified amino acids into libraries can be done by taking advantage of concurrent development in reassigning the 15 genetic code (Noren et al., 1989, Science 244:182-188; Benner, 1994, Trend. BioTech. 12:158-163) and the charging of specific tRNAs with the desired amino-acid (Cornish et al., 1994, Proc. Natl. Acad. Sci. USA 91:2910-2914). See also Ibba and Hennecke, 1994, Bio/Technology 12:678-682

20 (particularly Table I), and references cited therein. These

- pre-charged tRNAs are then utilized in the in vitro translation system to incorporate the non-standard amino acid into the library of choice. The position of incorporation can be either random (variant) or defined (invariant). The 25 defined case can be chosen to maximize the utility of the
 - resulting placement of the non-natural functional group to maximize either binding properties or the ability to perform structural measurements. Similar techniques may be used to incorporate non-standard amino acids into the peptides.
- In a specific embodiment, an iterative approach to library construction can be taken, as structural information on the mode of binding to a given target is obtained. For example, information from structural analysis can be used to make libraries with library members containing chemical
- 35 backbones that match known chemical scaffolds, enhance solubility or membrane permeability, reduce effect of water on structure, and incorporate other physical parameters

suggested by structural analysis. Use of algorithmically optimized library inserts can be used to increase the chances of finding binders of interest (see e.g., Arkin and Youvan, 1992, Bio/Technology 10:297-300).

In other embodiments, the following can be used to improve library use in both phage and bacterial systems: production of libraries in bacteria which overproduce the chaperonins GroES and GroEL (Soderlind et al., 1993, Bio/Technology 11:503-507), and production in E. coli strains which prevent degradation in the periplasmic space (Strauch and Beckwith, 1988, Proc. Natl. Acad. Sci. USA 85:1576-1580; Lipinska et al., 1989, J. Bacteriology 171:1574-1584). Purified cofactors such as GroES and GroEL could also be directly added to an in vitro expression and selection 15 system.

5.3. SCREENING OF DIVERSITY LIBRARIES

Once a suitable diversity library has been constructed (or otherwise obtained), the library is screened to identify 20 binders having binding affinity for the target. Screening is done by contacting the diversity library members with the target molecule under conditions conducive to binding and then identifying the member(s) which bind to the target molecule. Screening the libraries can be accomplished by any 25 of a variety of commonly known methods. See, e.g., the following references, which disclose screening of peptide libraries: Parmley and Smith, 1989, Adv. Exp. Med. Biol. 251:215-218; Scott and Smith, 1990, Science 249:386-390; Fowlkes et al., 1992; BioTechniques 13:422-427; Oldenburg et 30 al., 1992, Proc. Natl. Acad. Sci. USA 89:5393-5397; Yu et al., 1994, Cell 76:933-945; Staudt et al., 1988, Science 241:577-580; Bock et al., 1992, Nature 355:564-566; Tuerk et al., 1992, Proc. Natl. Acad. Sci. USA 89:6988-6992; Ellington et al., 1992, Nature 355:850-852; U.S. Patent No. 5,096,815, 35 U.S. Patent No. 5,223,409, and U.S. Patent No. 5,198,346, all to Ladner et al.; Rebar and Pabo, 1993, Science 263:671-673; and PCT Publication No. WO 94/18318. See also the references

```
cited in Section 5.2 hereinabove (disclosing libraries)
                                                                                                                   rding methods for screening. contacting the library by contacting the library harvesting the library contacting th
                                                                           Screening can be carried out by contacting harvesting examples of examples of the library members that bind to the target.

Members with members that bind to the target.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               Exambles of
                                                                                                those library members that bind to the target. Examples are that bind to the target. 1988. The target and smith 1988 are termed "panning" techniques and smith 1988. The such screening methods, termed in parmley and smith and smith termed are the such screening methods.
                                                          regarding methods for screening.
                                                                                                         such screening methods, termed "panning" techniques are gene in parmley and Smith, 13:422-427:

Such screening methods, termed "panning" techniques 13:422-427:

1992. BioTechniques 13:422-427:

1992. BioTechniques 13:422-427:
                                                                   members with an immobilized target molecule and to the target.

Sthose library members that tarmed "narring" rechni
                                                                                                                   described by way of example in parmley and smith, 13:422-427;

described by way of example 1992, and in references cited

73:305.318; Fowlkes et al., WO 94/18318; and in references cited
WO 96130849
                                                                                                                             13:305.318; Fowlkes et al., 1992, and in references to scre

NO. 94/18318; that can be used to scre

PCT Publication In panning methods the panning methods to screen the panning methods.

here nabove.
                                                                                                          PCT Publication No. WO 94/18318; that can be immobilized on per publication In Panning methods can be immobilized on the target molecule can be immobilized to some target molecule can be immobilized on the target molecule can be immobilized
                                                                                                                                                nereinabove. In panning methods that can be used to scree methods that can be immobilized on the target molecule beads. Secharose etc. the libraries such as magnetic beads.
                                                                                                                                                       the libraries, such as magnetic beads, embodiments, the plates, used in plates, used in beads, used in such as magnetic beads.
                                                                                                                                                                             beads used in columns. In particular embodiments, the has incorporated an effect.

beads used target molecule which can be used to effect.

immobilized target above. Which can be used to effect.
                                                                                                                                                           immobilized target molecule has incorporated an effect which can be used to earther which can be hinding narther the tag's binding narther the tag's binding have attaching the tag's hinding have attached an effect.
                                                                                                                                                                                                  tag, as described above, which can be used to effect to the tag, s binding partner to the tag, as described attaching the tag, s binding partner to the tag, as described attaching the tag, as described attaching the tag, as described attaching the tag, as described above, which can be used to effect to the
                                                                                                                                                                                                                                                                     red solid phase. the primary method of selecting from primary method of affinity capture of solid phase plastic affinity capture in one embodiment, solid phase plastic affinity capture in one embodiment, solid phase plastic affinity capture
                                                                                                                                                                                                                               In one embodiment, the primary method of selecting from the primary method of selecting capture of solid phase plastic affinity in the libraries the use of solid phase prior to its use in the libraries the rarget molecule prior to its use of solid phase plastic affinity in the libraries is the rarget molecule prior to its use in the libraries of solid phase plastic affinity in the libraries of solid phase planting in the libraries of solid 
                                                                                                                                                                                                              libraries is the targer molecule prior to its use improve to improve the targer molecule prior to its use improve to improve the targer molecule prior to its use improve to its use improve to its use improve to its use in the improve to its use in the improve to impossible the targer molecule prior to its use improve the improve
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          This method can be improved
                                                                                                                                                                                                                                                              selection (screening) process. selectivity and specificity.

This method can be improved

This method can be specificity.

This method can be specificity.

This method can be improved

This 
                                                                                                                                                                                                                                                                       upon to increase throughput, selectivity and specificity.

upon to increase throughput, can be replaced with magnetic supports can be large beads can be used to support solid phase plastic based systems.
                                                                                                                                                                                                                                                                                 Solid phase plastic supports systems; large due to steric particles. particles.
                                                                                                                                                                                                                                                              Particles. are not believed to be systems.

Particles are not be in bacterial systems.

particles for use in bacterial systems.
                                                                                                                                                                                                                                                        selection (screening) process.
                                                                                                                                                                                                                                                                                                             hindrance, to be avoided by using (2<0.5 mm) (Miltenyi et al., hindrance, with small particles separation with small particles
                                                                                                                                                                                                                                                                                                                           hindrance, or he similar his minimum hindrance, his he similar his minimum hindrance.
                                                                                                                                                                                                                                                                                                                                                                                                Cytometry 11:231-238). involving the use of a pertide involving the use of a pertide protein involving the use of a pertide involving the use of a pertial the use of a pertine use of a pertine use of a pertial the use of a pertine use of a pertine use of a pertine use of a pertial the use of a pertine use of a perturbation of a perturbat
                                                                                                                                                                                                                                                                                                                                                                  phage display library, selection of a binder protein now heir a bacteriophage it (the nww heir)

phage display the surface of a bacteriophage it (the nww heir)

expressed on protein and the nww that encodes it (the nww heir)
                                                                                                                                                                                                                                                                                                                                                      In a specific embodiment involving the use of a f
selection of a binder protein
selection of a bacterionhade thus selection
phage display the surface of a bacterionhade the surface of a final surface of a bacterionhade the surface of a bacterionh
                                                                                                                                                                                                                                                                                                                                                                           expressed on the surface of a bacteriophage thus selects both the surface of a bacteriophage it (the DNA the encodes it (the DNA that encodes it (the DNA the pollowing binding between the following binding between the binder protein and the phage particle).
                                                                                                                                                                                                                                                                                                                                                                                     the binder protein and the DNA that encodes it (the DNA being the binding between from Following binding released from the phage particle).

The binder protein and library members. The phage are released from the phage and library members.
                                                                                                                                                                                                                                                                                                                                       1990 Cytometry 11:231-238).
                                                                                                                                                                                                                                                                                                                                                                 within the phage particle). Following binding between the complex members, phage are released from members, phage and library members target molecule complex binder target molecule and which the binder target molecule are support on which the binder target molecule support of target molecules are support to the binder target molecule support of target molecules are support to the binder target molecules are supported to the binder target molecules are supp
                                                                                                                                                                                                                                                                                                                                                                                                          target molecule and library members, phage molecule complex binder target molecule and which the binder target by infection F con a solid support on which amplified.
                                                                                                                                                                                                                                                                                                                                                                                                                   a solid support on which the binder-target molecule complex infecting E. coli
```

and propagating each isolated binding phage. Repeating this process of affinity capture and amplification allows those peptides which bind with the highest affinity to the target molecule to be selectively enriched from the original.

5 library.

In one particular embodiment, presented by way of example but not limitation, a phage display library can be screened as follows using magnetic beads (see PCT Publication No. WO 94/18318):

10 Target molecules are conjugated to magnetic beads, according to the instructions of the manufacturers. The beads are incubated with excess bovine serum albumin (BSA), to block non-specific binding. The beads are then washed with numerous cycles of suspension in phosphate buffered saline (PBS) with 0.05% Tween® 20 and recovered by drawing a strong magnet along the sides of a plastic tube. The beads are then stored under refrigeration, until use.

An aliquot of a library is mixed with a sample 20 of resuspended beads, at 4°C for a time period in the range of 2-24 hrs. The magnetic beads are then recovered with a strong magnet and the liquid is removed by aspiration. The beads are then washed by resuspension in PBS with 0.05% Tween® 20, and 25 then drawing the beads to the tube wall with the magnet. The contents of the tube are removed and washing is repeated 5-10 additional times. 50 mM glycine-HCl (pH 2.0), 100 μ g/ml BSA solution is added to the washed beads to denature proteins and 30 release bound phage. After a short incubation, the beads are drawn to the side of the tubes with a strong magnet, and the liquid contents are then transferred to clean tubes. 1 M Tris-HCl (pH 7.5) 35 or 1 M NaH, PO, (pH 7) is added to the tubes to neutralize the pH of the phage sample. The phage are then diluted, e.g., 10⁻³ to 10⁻⁶, and aliquots

```
Plated With E. coli DHSQF cells to determine the
                                                                                                                                                                                                                Placed with E. coli DHSar, cells to determine to the sample.

Placed with E. coli DHSar, units of the sample.

The forming units of the sample.
                                                                                                                                                                                                                                                     certain cases, the platings are done in the discrimination of yGal and IPTG for color with the presence of yGal and IPTG for many and the presence of a series of the presence of the presence
                                                                                                                                                                                                                                    number or plaque torming units of the in the the the the platings are done in the certain cases, and rome for color discourse the platings are color discourse the certain of vortices and the certain of vortices are color discourse the certain of vortices are colored to the ce
                                                                                                                                                                                                                                                                         Presence of XGal and IPTG for color discrimination in the plaques are him in the plaques of the innit of plaques of plaques white
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          The titer.
                                                                                                                                                                                                                                                                                                                                      is also determined for another non-limiting example, as yet another non-limiting exam
WO 96130849
                                                                                                                                                                                                                                                        Alternatively as yet another non-limiting example, despectively as yet another phage expressing pertials (see pc)

Alternatively as yet another non-limiting example, despecting pertials another non-limiting example, despectively phage expressing pertials as yet another non-limiting example, despectively as yet another non-limiting example.
                                                                                                                                                                                                                screening a diversity library of phage expressing peptides (see pcr addressed by panning using microtiter plates (see pcr post of phage expression with the plates and panning using panning as follows:

10 can be achieved by 94/18318) as follows:
                                                                                                                                                                                                                                                                                                                                                                                                                                           The target molecule is diluted and a small adsorbed and a small is adsorbed and a small aliquot of target molecule solution is adsorbed and a small aliquot of target molecule aliquot of microtires aliquot aliquot of microtires aliquot aliquot of microtires aliquot of microtires aliquot of microtires aliquot of microtires aliquot a
                                                                                                                                                                                                                                                                                                 can be achieved by panning using microcite to the publication No. wo ga/la318) as follows:
                                                                                                                                                                                                                                                                                                                                                                                                                                                        aliquot of target molecule solution of RGA solution (1)

aliquot wells of microtiter plates of RGA solution (1)

onto wells at 40°C)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                            onto wells of microtiter plates of BSA solution the overnight at 4°C). An aliquot of added and the overnight at 100 mm Nawco. In a solution the overnight at 100 mm Nawco.
                                                                                                                                                          5
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             overnight at 4°C). An aliquot of BSA solution the maliquot of BSA solution the management of the sadded and the sadde
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 mg/ml, in 100 mm NaHCO; ph 8.51 15 added and to the incubated at room temperature for incubated at room plate
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     plate incubated at room temperature for 1 hr. out

plate incubated at room temperature are flicked out

contents of the microtiter plate are fully with pre-noce
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      contents of the microtiter place are flicked of the microtiter place are flicked of the microtiter place are fully with passion of the microtiter place are fully with passion of the microtiter place are fully with passion of the microtiter place are flicked of the microtiter place 
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              and the wells washed carefully with PBS-0.05% free repeatedly washed from the Plates are repeatedly aliming of Tween 20.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  Tween 20. The Plates are repeatedly washed free and the molecules.

The Plates are repeatedly washed free a small aliquot of and the molecules.

The Plates are repeatedly washed free are molecules.

The Plates are repeatedly washed free free are molecules.

The Plates are repeatedly washed free free are more and the molecules.

The Plates are repeatedly washed free free are more are more
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   of unbound target molecules. A small aliquot of the into each well and the for 2-24 remnerarity for 2-24 remnerarity for phage solution is intimared at room remnerarity for 2-24 remnerarity for 2-24
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          phage solution is introduced into each well and the phage solution is introduced temperature for 2-24.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             are incubated at room temperature for 2-24

The contents
                                                                                                                                                                                                                                                                                                                                         25
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   out and washed repeatedly. The plates are for 20 incubated with wash solution to allow hours of the incubated with the remnerature to allow hours of the incubated with the remnerature to allow hours of the incubated with the remnerature to allow hours of the incubated with the remnerature to allow hours of the incubated with the remnerature to allow hours of the incubated with the incubated
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   nrs. and washed repeatedly in in and income out and washed with wash solvering in a second income of the places are
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    incubated with wash solution in each well for 20 hound phage to allow bound phage minutes at room temperature to allow he relative to minutes at room temperature to minutes at room temperature to allow he relative to al
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        minutes at room temperature to allow bound phage

minutes at room temperature

minutes at room temperature

with rapid dissociation constants

minutes at room temperature

with rapid are reasonable and the reasonable are reasonable at reasonable and the reasonable are reasonable at reaso
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             with rapid dissociation constants to be released.

With rapid dissociation washed five more times to remove

The wells are

The wells are
                                                                                                                                                                                                                                                                                                                                                                                                                                       20
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   inbound phage the phage bound to the wells, a phage bound to the wells, and to recover the phage bound the phage bound to recover the phage bound the phage bound the phage bound to recover the phage bound to recover the phage bound to recover the phage bound the phage boun
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         To recover the phage bound to the weller a pi

change is used.

change is and malana sea solution is added to the

change is used.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 change 15 used. An aliquot of 50 mm glycine KCl

the

change 15 used. An aliquot of 50 mm glycine the

(ph 2.0) 100 to densitive arotein and relative arotein and relative arotein arotein and relative arotein arotei
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              (PH 2.0), 100 µg/ml BSA BOLUTION 15 and release bound washed wells to denature proteins and release hours washed wells to denature proteins and release hours are continuous and release bound.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      Wells to denature Proteins and release Dound

After 10 minutes at 65°C, the contents

After 10 minutes at 10 minut
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            25
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           all unbound phage.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       pnage. Atter 10 marnutes at book the contents in the contents 
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           then transferred into clean tubes, and a small (pH aliquot of 1 M Tris-HCl (pH 7.5) or 1M NaH, PO,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              30
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               35
```

PCT/US96/04229

5

10

7) is added to neutralize the pH of the phage sample. The phage are then diluted, e.g., 10⁻³ to 10⁻⁶ and aliquots plated with E. coli DH5αF' cells to determine the number of the plaque forming units of the sample. In certain cases, the platings are done in the presence of XGal and IPTG for color discrimination of plaques (i.e., lacZ+ plaques are blue, lacZ- plaques are white). The titer of the input samples is also determined for comparison (dilutions are generally 10⁻⁶ to 10⁻⁹).

By way of another example, diversity libraries expressing peptides as a surface protein of either a particle or a host cell, e.g., phage or bacterial cell, can be screened by passing a solution of the library over a column 15 of the target molecule immobilized to a solid matrix, such as sepharose, silica, etc., and recovering those particles or host cells that bind to the column after washing and elution.

In yet another embodiment, screening a library can be performed by using a method comprising a first "enrichment" 20 step and a second filter lift step as described in PCT Publication No. WO 94/18318.

Several rounds of serial screening are preferably conducted. In a particularly preferred aspect, each round is varied slightly, e.g., by changing the solid phase on which 25 immobilization occurs, or by changing the method of immobilization on (e.g., by changing the linker to) the solid phase. When using a phage display library, the recovered cells are then preferably plated at a low density to yield isolated colonies for individual analysis. By way of 30 example, the following is done: The individual colonies are selected, grown and used to inoculate LB culture medium containing ampicillin. After overnight culture at 37°C, the cultures are then spun down by centrifugation. Individual cell aliquots are then retested for binding to the target 35 molecule attached to the beads. Binding to other beads, having attached thereto a non-relevant molecule, can be used as a negative control.

In a specific embodiment, different rounds of screening can respectively involve selection against targets in primarily their purified form, and then in their natural state (e.g., on the surface of a mammalian cell) (see, e.g., 5 Marks et al., 1993, Bio/Technology 11:1145-1149, describing selection against cell surface blood group antigens).

In other examples, subsequent rounds of screening can involve immobilization of the target molecule by attachment at different ends (e.g., amino or carboxy-terminus) of the 10 target molecule to a solid support, or presentation of library members by attachment to or fusion at different ends of the library members.

By way of other examples of screening methods that can be used, genetic selection methods can be adapted for 15 screening of libraries, or can be used in a recursive scheme. Thus, in a specific aspect, the invention provides screening methods in which methods allowing high throughput and diversity screening (e.g., screening phage display or polysome libraries against a ligand) are utilized in initial 20 rounds, with subsequent rounds employing a genetic selection technique, in which the presence of a binder of appropriate specificity increases the activity of or activation of a transcriptional promoter or origin of replication. Genetic selection techniques that can be adapted for use (e.g., by25 inserting random oligonucleotides in the test plasmid) include the two-hybrid system for selecting interacting proteins in yeast, replicative based systems in mammalian cells, and others (see, e.g., Fields & Song, 1989, Nature 340:246-246; Chien et al., 1991, Proc. Natl. Acad. Sci. USA 30 88:9578-9582; Vasavada et al., 1991, Proc. Natl. Acad. Sci. USA 88:10686-10690). Thus, in a specific embodiment, compounds are produced as fusion proteins, and contacted with a different fusion protein comprising a target fused to another molecule, in which specific binding of the fusion 35 proteins to each other results in an increase in activity or activation of a transcriptional promoter or an origin of replication. In a specific embodiment, a genetic selection

method is used in a later round of screening to either select directly for a library member that binds to a target molecule, or to select a library member that competitively inhibits binding of a ligand to the target molecule.

Several exemplary methods for screening a phage/phagemid library are presented by way of example in Section 6.4 hereinbelow. An exemplary method for screening a polysomebased library is presented in Section 6.3.3 hereinbelow.

Once binders are selected from a diversity library which

10 bind to a target molecule of interest, additional assays are
preferably, although optionally, performed, including but not
limited to those described below. Thus, in vivo or in vitro
assays can be performed to test whether binding of a binder
to the target molecule affects the target molecule's

- 15 biological activity; binders that exert such an effect are preferred for use in subsequent steps of the invention.

 Alternatively, or in addition, competitive binding assays can be carried out to test whether the binder competes with other binders or with a natural ligand of the target molecule, for
- 20 binding to the target molecule; binders that compete with each other, and that compete with the natural ligand, are preferably selected for use in subsequent steps of the invention. Alternatively, or in addition to the above assays, the binding affinity of binders for the target
- 25 molecule is determined, by standard methods, or by way of example, as described in Section 6.5 infra. Binders of the highest affinity are preferred for use in subsequent steps of the invention.

30 5.4. DETERMINING THE SEQUENCE OR CHEMICAL FORMULA OF BINDERS

Many of the references cited in Section 5.2 and 5.3 hereinabove, which disclose library construction and/or screening, also disclose methods that can be used to determine the sequence or chemical formula of binders isolated from such libraries. By way of example, a nucleic acid which expresses a binder can be identified and recovered

from a peptide expression library or from a polysome-based library, and then sequenced to determine its nucleotide sequence and hence the deduced amino acid sequence that mediates binding. (In an instance wherein the sequence of an 5 RNA is desired, cDNA is preferably made and sequenced.) Alternatively, the amino acid sequence of a binder can be determined by direct determination of the amino acid sequence of a peptide selected from a peptide library containing chemically synthesized peptides. In a less preferred aspect, 10 direct amino acid sequencing of a binder selected from a peptide expression library can also be performed.

Nucleotide sequence analysis can be carried out by any method known in the art, including but not limited to the method of Maxam and Gilbert (1980, Meth. Enzymol. 65:499-15 560), the Sanger dideoxy method (Sanger et al., 1977, Proc. Natl, Acad. Sci. U.S.A. 74:5463), the use of T7 DNA polymerase (Tabor and Richardson, U.S. Patent No. 4,795,699; Sequenase[™], U.S. Biochemical Corp.), or Taq polymerase, or use of an automated DNA sequenator (e.g., Applied Biosystems, 20 Foster City, CA).

Direct determination of the chemical formulas of nonpeptide or peptide binders can be carried out by methods well known in the art, including but not limited to mass spectrometry, NMR, infrared analysis, etc.

- In preferred aspects involving certain types of libraries well known in the art, sequencing or the use of known analytic techniques for chemical formula determination will not be necessary. In some such libraries, the identity and composition of each member of the library is uniquely
- 30 specified by a label or "tag" which is physically associated with it and hence the compositions of those members that bind to a given target are specified directly (see, e.g., Ohlmeyer et al., 1993, Proc. Natl. Acad. Sci. USA 90:10922-10926; Brenner et al., 1992, Proc. Natl. Acad. Sci. USA
- 35 89:5381-5383; Lerner et al., PCT Publication No.
 WO 93/20242). In other examples of such libraries, the
 library members are created by step wise synthesis protocols

accompanied by complex record keeping, complex mixtures are screened, and deconvolution methods are used to elucidate which individual members were in the sets that had binding activity, and hence which synthesis steps produced the 5 members and the composition of individual members (see, e.g., Erb et al., 1994, Proc. Natl. Acad. Sci. USA 91:11422-11426).

Step 2 of the invention provides as output N binding library members (binders) and their sequences or chemical formulas.

10

5.5. CANDIDATE PHARMACOPHORE SELECTION

The prior diversity library screening, step 2, determines a set of size N of specifically binding members from one or more diversity libraries. While the binders are 15 preferably but not necessarily isolated from one or more diversity libraries (e.g., binders need not be isolated from diversity libraries; known binders can be simply provided), the following description shall refer to the preferred embodiment wherein diversity library members are the binders. 20 It will be apparent that the description is also readily applicable to binders that are not isolated from diversity libraries.

The pharmacophore responsible for the library member binding is preferably determined by an overall select and 25 test method in this and subsequent steps. In general, a pharmacophore is specified by the precise electronic properties on the surface of the binder that causes binding to the surface of the target molecule. In the preferred embodiment, these properties are specified by the underlying, 30 causative, chemical structures. Chemical structures are specified generally by groups such as -CH₂-, -COOH, and -CONH₂. The preferred pharmacophore representation consists of a specification of the underlying chemical groups and their geometric relations. The more precisely the geometric relations are specified, the more preferred. In preferred but not limiting aspects, the geometric relations are precise to at least 0.50 Å, and most preferably, at least 0.25 Å. A

pharmacophore will usually comprise 2 to 4 of such groups, with 3 being typical. However, for complex protein recognition targets, a pharmacophore may comprise a greater number of groups. For example, it is possible that the 5 entire 6 amino acid sequence, -X_i-, may be needed for a member of the preferred CX_iC library to bind to complex targets, in which case the pharmacophore includes the entire binder.

Considering by way of example, the case of binders isolated from the preferred library, of sequence CX₆C, the 10 chemical groups defining a peptide pharmacophore are terminal groups on amino acid side chains. Typically, therefore, a sequence of two to four contiguous amino acids will contain the pharmacophore of interest. For example, Fig. 11 illustrates an Arginine-Glycine-Aspartate sequence forming a 15 well known platelet aggregation inhibiting pharmacophore, which is defined by the positions and orientations of the adjacent -CN₃H₄, -CαH₂-, and -COOH groups. Pharmacophores formed by discontiguous amino acids are not likely to occur in the preferred library due to the conformational constraint 20 on the short peptide imposed by the disulfide bridge.

The selection step determines candidate amino acid sequences in each binder that define a candidate pharmacophore by the positions of their terminal groups. Candidate selection depends substantially only on the 25 chemical structures of the amino acid side chains and terminal groups (only very rarely on backbone groups). Geometric structure is not yet available and cannot be used for candidate selection. In the preferred embodiment, amino acids are grouped into homologous groups defined by group 30 members having similar side chain structure and activity (see infra). Candidate pharmacophores are found by searching the sequences of the N binders for short sequences of homologous amino acids. This search will produce at least one candidate, because all the binders share the actual 35 pharmacophore. Several candidates will usually be found since geometric information is ignored, and the search is thereby underdetermined.

Fig. 2A illustrates an exemplary method of performing the search for homologous sequences. Although this method is illustrated as searching for homologous contiguous sequences of length 3, it is easily adaptable to search for homologies 5 of other lengths and also for discontiguous homologous sequences. If no candidate pharmacophores of length 3 have a consistent consensus structure, then pharmacophores of length 2, 4, or longer or discontiguous sequences must be searched and selected for test. For some complex targets, the 10 pharmacophore may include the entire variable part of the library member. The exemplary method is a simple depth-first search for matching amino acid strings. More sophisticated string search methods are known and are equally applicable to this invention.

The method begins with the administrative steps 201 and 15 202 of labeling the binders with integers from 1 to N and assigning the string variable 'ABC' to the next left most sequence of three amino acids to test in binder 1. If this is the first candidate selection, 'ABC' will be at the left 20 most position in binder 1. If prior candidates have been selected, 'ABC' will be assigned one amino acid to the right of its prior assignment. The FOR loop, formed by steps 203, 206, and 207, then selects each binder from 2 to N for scanning for a sequence homologous to 'ABC'. Step 203 does 25 loop administration. Step 206 does the scanning. homologous sequences are found, test 207 loops back to scan the next binder. If homologous sequences have been found in all binders from 2 to N, the loop exits at step 204. In this case 'ABC' is a string in binder 1 which is homologous to 30 other strings in all remaining binders and is thus a candidate pharmacophore. The method exits at 205 for this candidate to be structured and tested for whether it is the actual pharmacophore. If a binder does not have a sequence homologous to 'ABC', then this string is not a candidate. In 35 this case, test 208 determines if 'ABC' is at the right end of binder 1. If so, there are no more homologies to test for

and the method exits at 209. If not, then 'ABC' is advanced

one amino acid to the right 210 and the scan of all binders is repeated beginning at 203.

Fig. 2B illustrates how string variable 'ABC' is scanned across binder 1, represented schematically by 220. First, 5 'ABC' is assigned to X₁X₂X₃ at 221, then to X₂X₃X₄ at 222, to X₁X₄X₅ at 223, and finally to X₄X₅X₆ at 224.

Given an assignment to 'ABC', step 206 scans each other binder, for example binder K with K>1, for homologous sequences. This is simply done by comparing all contiguous 10 substrings of binder K with 'ABC' to determine if they are homologous. They are homologous if corresponding amino acids in the substring and 'ABC' are homologous. In turn, two amino acids are homologous if they satisfy established homology rules. Each homologous sequence found in binder K 15 defines a separate candidate pharmacophore, if sequences homologous to 'ABC' are found in all other binders.

In a case where discontiguous homologous sequences are sought, 'ABC' is assigned to amino acids in discontiguous positions in binder 1 and then compared for homologies to 20 amino acids in the same relative positions throughout the other binders.

Various rules of amino acid homology may be used in this invention. In the preferred embodiment, amino acids are homologous if they are found in the same class of amino 25 acids, based on side chain activity (see Lehninger, Principles of Biochemistry, (1982), chap. 5). Preferred homologous groups of amino acids are as follows. The nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and 30 methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and 35 glutamic acid. The foregoing classes may be modified by those skilled in chemical arts to create finer classifications. For example, phenylalanine and tryptophan

could be placed in a separate aromatic nonpolar group. Further, homology rules could depend on amino acid sequence, such as by dividing contiguous doublets or triplets of amino acids into homology groups.

5 The invention is not limited to the above-described exemplary method of selecting candidate pharmacophores. Any automatic method of selecting candidates that depends only on chemical structure of binder library members, preferably expressed in terms of building block composition and 10 sequence, can be used. For example, in the case of the preferred CX₆C library, candidates could be selected by a clustering analysis performed on the entire amino acid string

in a multi-dimensional space.

This above method of selecting candidate pharmacophores 15 is not limited to the preferred CX,C diversity library. example, this method is immediately applicable to any diversity library having members comprising building blocks linked by a linear backbone by simply specifying rules of homology appropriate for the building blocks. These homology 20 rules would group building blocks presenting similar structure and reactivity to targets. This method then selects candidates comprising sequences of homologous building blocks present on all the binding library members. If the library members do not have a linear backbone, a 25 related candidate selection method can be used. case, the search for homologous building blocks would need to be confined to adjacent building blocks. Adjacent building blocks in this case are those building blocks brought physically close by whatever chemical structures form the 30 library members (instead of simply being linearly adjacent on a backbone). An adjacency determination would be specific to the particular chemical structure and would be algorithmicly specified. In addition appropriate rules of homology would be specified. The method would then select candidates 35 comprising groups of adjacent, homologous building blocks, a group being present on each binding library member.

The above-described step is the selection step of the overall select and test method. Distance measurements and Monte Carlo structuring, steps 4 and 5, determine a consensus pharmacophore structure for the candidate, if possible. If a consensus is found, the candidate is the actual pharmacophore. If a consensus is not found, this selection step must be revisited, and a new candidate selected for test.

10 5.6. <u>INTRAMOLECULAR DISTANCE MEASUREMENTS</u>

Having obtained N binders, their chemical building block structures (chemical formula or primary sequence), and the identification of a candidate pharmacophore in each binder, steps 4 and 5 of the method of this invention cooperatively 15 determine a precise spatial structure for the candidate pharmacophore (if it exists; if not, a new candidate pharmacophore is selected.) In the preferred (but not limiting) embodiment of this invention, N members of the CX₆C library that specifically bind to the protein target of 20 interest have been screened; their sequences determined; and a candidate pharmacophore consisting of homologous triplets (more generally from 2 to 6 mers) of amino acids has been determined in each binder.

Step 4 measures one or more strategic distances,

25 preferably no more than 10-20, e.g., 1-10 or, more
preferably, 1-5 interatomic distances are measured. The
remainder of the structure is determined in subsequent steps,
other than by direct measurement. The interatomic distances
measured in step 4 are preferably with an accuracy of at

30 least 2 Å, more preferably at least 1 Å or 0.5 Å or 0.25 Å,
and most preferably at least 0.05 Å. Thus, in a preferred
but not limiting embodiment, distances in the pharmacophore
are specified to at least approximately 0.25 Å. Step 5,
using the CCMBC computational method, then completes

35 determination of the pharmacophore structure at a high
resolution and the structures of the rest of the binder
molecules with a secondary resolution. Having a high

resolution structure for the pharmacophore of interest is orders of magnitude more useful than having a low resolution structure for an entire binder. Consequently, steps 4 and 5 focus resources on the former problem.

- A distance measurement method is preferred for use if it meets certain conditions, as follows. First, accuracy of distance measurements is preferably better than at least 0.25 Å for distances on the order of those between amino acids in a peptide. Second, measurement conditions preferably
- 10 approximate target binding conditions, i.e., are approximately physiologic. For example, crystallization, which may induce conformational changes, is preferably avoided. Also, the employed measurement methods preferably allow one binder sample to be measured when dry, when
- 15 hydrated and when bound to the target molecule of interest, thereby observing the effects of water and conformational changes on binding. Third, the measurement method is preferably quick and inexpensive.
- Important advantages are conveyed by these certain

 20 conditions. First, as the method of the invention determines high resolution pharmacophore structures, use of distances less accurate than the intended results would almost certainly result in decreased resolution. Second, as the CCMBC structure determination method approximates the
- 25 structural effects of hydration and target binding, use of accurate distances including the physical effects of hydration or binding helps increase the resolution of the computational results. These distances as used in the CCMBC method pull the binder structures towards a more accurate
- 30 representation both of the bound, hydrated pharmacophore and also of the remainder of the binder molecule without a computationally burdensome inclusion of water molecules and without knowledge of the target molecule's structure.

REDOR NMR is the preferred method of distance

35 determination. REDOR is a solid phase NMR technique which directly measures the inter-nuclear dipole-dipole interaction strength between two spin % nuclear species, denoted D_{AB} where

A and B are the two nuclear species measured. The internuclear distance between A and B is simply determined from D_{AB} by the following equation:

$$D_{AB} = \frac{h \gamma_A \gamma_B}{2\pi R_{AB}^3} \tag{1}$$

where R_{AB} is the inter-nuclear distance, h is Planck's constant, and γ_A , and γ_B are the respective gyromagnetic 10 ratios of nuclei A and B. REDOR is typically accurate to less than 0.05 Å and can generally measure distances up to about 8 Å.

Any two nuclear species observable and resolvable by NMR methods and, preferably, adaptable to chemical inclusion in the diversity library members of interest, may be the basis of REDOR measurements. Although the subsequent description is often directed to distance determinations between ''C and ''N nuclei in members of a preferred library comprising the sequence CX₆C, this invention is not so limited. One skilled in the art can readily adapt the method for use in making measurements of other types of molecules (e.g., peptides and nonpeptides); additionally, other nuclear species may be used. Other common spin % species that can be used include but are not limited to ''P and the halogen 'F.

General references on NMR techniques are Slichter,

Principles of Magnetic Resonance, Berlin, Springer-Verlag,

(1989) and Mehring, <u>High Resolution NMR in Solids</u>, Berlin,

Springer-Verlag (1983). REDOR references include Gullion et
al., <u>Rotational-echo double-resonance NMR</u>, J. Magn. Res.

30 81:196-200 (1989); Pan et al., <u>Determination of C-N</u>
internuclear distance by rotational-echo double-resonance NMR
of solids, J. Magn. Res. 90:330-40 (1990); Garbow et al.,
Determination of the molecular conformation of melanostatin
using 13C, 15N-REDOR NMR spectroscopy, J. Am. Chem. Soc.

35 115:238-44 (1993), all of which are incorporated herein by reference.

Other solid phase NMR techniques are applicable but less preferred. These include but are not limited to those disclosed in Kolbert et al., Measurement of internuclear distances by switched angle spinning, J. Physical Chemistry 58:7936 et seq. (1994), and in Raleigh et al., Rotational Resonance NMR, Chemical Physics Letters 146:71 (1988). These techniques measure homonuclear distances only to 0.5 Å accuracy and are less accurate than REDOR. Liquid phase NMR techniques of NOE (nuclear overhausser) and COESY

10 (correlation enhanced spectroscopy) can also be used but are less preferred. They require complex interpretation to

less preferred. They require complex interpretation to obtain comparable distance accuracy greater than 0.5 Å in small molecules with complete rotational freedom.

X-ray crystallography can also be used, although it is 15 much less preferred, since crystallization may induce conformational changes in the binder, and since binding to the target molecule may be necessary for crystallization.

In the case of REDOR measurements of the heteronuclear distances between ¹³C and ¹⁵N, ¹³C and ¹⁵N are introduced

20 ("labeled") at the positions between which a distance measurement is needed. The preferred embodiment of the invention measures the ¹⁵N NMR resonance. Since nearly all the ¹⁵N signal will originate with nuclear labels, very little background signal due to natural abundance nuclei need be accounted for. Alternatively, the ¹³C resonance may be measured, in which case the natural abundance background is subtracted from the measurements.

Since REDOR depends on observing the internuclear dipole-dipole interaction, the binder being measured should 30 be substantially stationary on the time scale of the NMR signal. The measurement system preferably ensures this condition. The substrate holding the binder to be measured can be chosen so as to restrain binder motion, or the measured sample may be cooled to restrain motion, or, 35 alternatively, the binder may be bound to its target molecule in order to restrain its motion.

Further details of the REDOR distance measurements will make reference to Fig. 3. This illustrates the measurement method for one labeling of one binder, which is repeated if the binder requires multiple labelings and also is repeated 5 for each binder. Subsequent description will focus on only one binder.

Step 41 chooses a binder labeling. Labeling is preferably done to obtain the most information about the pharmacophore consistent with chemical labeling opportunities 10 and available labeled amino acids. Backbone labeling, for example, labels the amide N of one amino acid and one of the backbone C's of a next adjacent or more distant amino acid. Backbone labeling is typically done in the backbone in the vicinity of the candidate pharmacophore. It might also be 15 done away from a candidate pharmacophore to confirm a previously determined structure as described for step 6. Side chain labeling strategies vary with the chemical opportunities offered by the candidate pharmacophore. If a terminal N is available, an adjacent side chain or backbone C 20 can be labeled. If not, the side chain C and backbone amino N can be labeled. Side chain labeling is preferably on side chains in the candidate pharmacophore. Preferred labeling in the candidate pharmacophore is either a backbone amino N and a nearby backbone C or a side chain C or, if available, a 25 side chain amino N and an adjacent or nearby side chain C.

In an alternative embodiment, to get the most structural information on the binders, these labelings are designed to select the actual major conformation from known possible conformations. For example, if it is known from preliminary determinations that a binder may exist in one of a few, e.g. two, major backbone or side chain folding patterns, the labelings are chosen to distinguish these conformations. Nuclear pairs labeled for measurement are preferably those that have significantly different distances in the possible conformations.

Multiple labeling of one binder to determine multiple distances at once is possible, for example, by including one

¹³C and several ¹⁵N nuclei, or vice versa, in one labeled molecule. Multiple labeling is limited, however, as is obvious to one skilled in the NMR arts, by chemical shifts of the various nuclear resonances. REDOR measurement of multiple ¹⁵N-¹³C distances requires that each spectroscopically observed ¹⁵N or ¹³C resonance have a distinguishable chemical shift. If these conditions are not met, several separately labeled versions of the binder are prepared and measured, one for each internuclear distance sought.

- Step 42 synthesizes the labeled binder after a labeling has been determined by applying these preferences and rules. In an embodiment wherein the binder is a peptide, variously labeled '1°C or '5N labeled amino acid reagents for the synthesis of the labeled binder are widely available from 15 commercial sources. A preferred supplier is Isotec Inc. (Miamisburg, OH). Other commercial sources include MSD Isotopes (Montreal, Canada) and Sigma Chemical Co. (St. Louis, MO). Step 42 has three substeps: linear peptide synthesis 43, cyclization 44 (by forming the disulfide bond), 20 and deprotection of the side groups 45. Synthesis and side chain deprotection are performed by solid phase peptide synthesis using standard Boc (tert-butoxycarbonyl) and Fmoc
- references for this method are Merrifield, J. Amer. Chem.

 25 Soc., vol 85, pp 2149 et seq. (1963); Caprino et al., J.

 Amer. Chem. Soc. (1970); and Stewart et al., Solid Phase

 Peptide Synthesis, Berlin, Springer-Verlag (1984), which are herein incorporated by reference. Cyclization is by conventional mild oxidation, well known in the chemical arts.
- 30 The method of these steps is detailed in Example 2 supra.

(9-fluorenylmethyloxycarbonyl) chemistry. Exemplary

To obtain accurate REDOR NMR measurements, the binder sample is preferably highly purified. Accordingly, it is preferable that the sample be at least 90% pure (but not necessary if spurious NMR signals can be discriminated), and 35 even more preferable that the sample be at least 95% pure. Such pure samples can be obtained as follows. In a first synthesis method, the binder peptide is synthesized directly

on the substrate to be used in the subsequent NMR measurements. In this case particular care is preferably taken with the standard solid phase synthesis steps of Example 2. By way of example, synthesis reagents should be 5 pure, adequate time should be allowed for diffusion of reagents and solvents throughout the interstices of the substrate resin, and between steps, prior reagents should be thoroughly washed from the resin before new reagents applied. That the purity, reaction time, and washings are adequate is 10 gauged by subsequent analysis. An aliquot of the resulting peptide-resin is taken, the peptide is cleaved (Example 2) and its purity analyzed by mass spectroscopy or high performance liquid chromatography (HPLC).

In a second synthesis method, the peptide can be

15 synthesized on any convenient solid phase substrate in a

standard manner and then cleaved from the substrate. The

peptide is purified by standard methods (e.g., HPLC) and then

attached to the NMR measurement substrate. The attachment

can be done by any methods known in the art, preferably at

20 either the amino- or carboxy-terminus, e.g., by condensation

of the free carboxy terminal group on the peptide with an

amino labeled resin, with the attachment step preceding

deprotection of any side chain carboxy groups on the peptide;

by use of heterofunctional linker groups, etc.

Great care is preferably exercised in forming the binder-substrate used for the REDOR NMR measurements. This invention is also directed to binder-substrates suitable to precise REDOR NMR measurements in the following environmental conditions: dry unbound, hydrated unbound, and bound to its molecular target molecule (e.g., in lyophilized or hydrated forms).

For any binder and any NMR measurement substrate utilized, the substrate should restrain the attached binder sufficiently so that binder motion will not average out the dipole-dipole interactions necessary for the REDOR measurement. Generally, this requires that the frequency of motion of the binder be less than the frequency of the

dipole-dipole interaction being observed, which varies with the nuclear species being observed and the measurement distance. For 13C-15N observations to 2.5 Å the binder motion frequency should be less than approximately 200 Hz; for 5 observations to 5 Å, less than approximately 30-50 Hz; and for observations beyond 5 Å, less than approximately down to 10 Hz. The more polar the substrate, such as glass beads or p-MethylBenzhydrilamine ["mBHA"] resin, the more are polar attached binders (such as are many peptides) restrained. 10 Less polar substrates, such as polystyrene resin, provide less restraints for polar binders. In an embodiment wherein a peptide comprising the sequence CX,C is bound to an mBHA resin with an glycine residue serving as a linker to a binding site on the resin, probably no additional steps need 15 be taken for 2.5 Å measurements. Additional steps that can be used, if needed, to slow binder motions include cooling the measurement sample to, for example, liquid N2 temperatures (approximately 77 °K) or binding to a large, relatively immobile target molecule.

Second, the net binder density is important and 20 typically is adjusted. The substrate preferably has an adjustable number of binder synthesis sites or binding sites per unit of substrate surface area. Too high a binder density on the substrate surface will cause inter-molecular 25 nuclear dipole-dipole interactions to distort the REDOR distance measurements. To obtain accurate intra-molecular distances, the peptides should be kept sufficiently far apart so that only intra-molecular nuclear dipole-dipole interactions are significant. Inter-molecular nuclear 30 dipole-dipole interactions are preferably kept less than about 10% of the intra-molecular interaction. In the case of ¹³C-¹⁵N measurements, this criterium can be monitored by observing 13C-13C dipolar couplings. As the dipole interaction falls off as R-3, keeping adjacent binders apart by more than 35 approximately 2-3 times the distance to be measured is sufficient. For measurements to 5 Å, this criterion can be

satisfied by keeping binders approximately 10 Å or more

apart. At a 10 Å spacing interfering ¹³C or ¹⁵N signals will not exceed 2.8 hz, which is sufficient attenuation for 30 hz or greater measurements.

In an embodiment wherein the binder is a peptide 5 comprising the sequence CX,C, that is synthesized on an mBHA resin that is also to serve as the NMR substrate, there is an additional upper bound on the peptide density. To prevent disulfide dimer formation in more than approximately 5% of peptides, the peptides are preferably kept apart by at least 10 their average size. Dimer formation and incorrect disulfide scaffolds result in unconstrained, flexible peptides of altered structure distorting the REDOR distance determination of the properly conformationally constrained, cyclized binder peptides. A 10 Å or more separation will meet this 15 requirement. In this case, more than 95% of the disulfide bonds will result in intended intra-molecular constraints. This separation may be adjusted based on a determination of actual dimer formation by chromatographic (e.g., HPLC) or mass spectroscopic analysis of the peptide after cleavage

NMR instrumental sensitivity places a lower bound on binder density. By way of example, for an adequate observed signal to noise ratio using a preferred NMR spectrometer, no less than approximately 10¹⁸ observed nuclear spins should be present in a 0.1 g sample. This translates to having a binder density of no less than approximately 0.017 mmole/g (1 mmole = 10⁻³ mole). For alternative NMR spectrometers with higher field magnets (³H Larmor frequency of 500 mHz), the

20 from the substrate (see Section 6.6, infra).

binder density may be as low as 0.0017 mmole/g.

30 A third substrate condition to be considered is pore size, which is relevant when measurement of binder bound to a target molecule is desired. In a preferred method of conducting such bound measurements, the substrate must have sufficient pore size so that the target molecules can diffuse 35 to all binders on the surface of the substrate and bind to them. For example, folded, moderate sized protein targets of 50 kd are typically roughly spherical with diameters of

PCT/US96/04229

approximately 50 Å. Preferable substrate pore sizes for use with such moderate sized protein targets are no less than 100-200 Å. Excessive pore sizes can result in a too dilute binder that decreases NMR signal intensity. The preferable 5 pore sizes also facilitate high purity peptide synthesis directly onto substrate resins by similarly facilitating diffusion of reagents and solvents to synthesis sites. Also, binder substrate binding is preferably of such a nature that it will not be disrupted under either dry conditions, aqueous 10 conditions, and conditions suitable to binder-target binding. Generally, adequate pore sizes are in the range of 100-500 Å, although this will vary with the size of the target molecule.

Solid phase substrates that can be used include but are not limited to mBHA resins, divinylbenzyl polystyrene resins, and glass beads. All of these substances can be manufactured to have binding sites in the range from 0 to 1.0 mmol/g. In addition, these substrates can be made so as to have the following surface areas: for mBHA about 100 m²/g, for polystyrene from 50-100 m²/g, and for glass from 0.1-100 m²/g. These substrates also can be manufactured so as to have a surface binding site density in the range of from 0 to 1.0 mmol/m². More generally any microporous material with a surface density of binding sites adjustable from 0 to at least 1.0 mmol/m², and preferably with pore sizes in the

25 preferred ranges, can be used. Suppliers of such adjustable resins include Chiron Mimotope Peptide Systems (San Diego, CA) and Nova Biochem (San Diego, CA).

Peptide binders can be synthesized directly on the surface of the substrates, by way of example as set forth in 30 Section 6.6 infra, to achieve a purity of preferably at least 90%, more preferably at least 95%. In the case of a peptide comprising the sequence CX₆C, the preferred peptide spacing on the substrate is no closer than approximately 10 Å, or a peptide density of no greater than one peptide every 100 Å².

35 Peptide synthesis on the preferred resin p-MethylBenzhydrilamine ["mBHA"] with 0.16 mmole/g of peptide binding sites, a surface of 100 m²/g, and a preferable pore

size of 100-200 Å results in a binder-substrate having such a preferable peptide surface density and suitable for accurate REDOR NMR measurements in dry, hydrated, and bound conditions. The total binder density is more than tenfold above instrumental sensitivity. The glycine linker provides a sufficient spacer from the substrate surface.

Steps 43, 44, and 45 in the preferred embodiment of the invention are carried out by one of a number of commercial peptide synthesis sources, such as Chiron Mimotope Peptide

10 Systems (San Diego, CA) and Nova BioChem (San Diego, CA).

Methods that can be used in these steps are known in the art. However, the preferred practice of these steps is detailed in the example in Section 6.6.

The invention thus provides a method of performing solid 15 state NMR, preferably REDOR NMR, measurements of molecules on a solid phase substrate. In one embodiment, the molecule is a compound having conformational degrees of freedom at the temperature of interest that are limited to torsional rotations about bonds between otherwise rigid subunits, the 20 torsional rotations respecting any conformational constraints. The molecule is preferably a peptide, more preferably a peptide of constrained conformation, and is most preferably a peptide having one or more cystines (e.g., comprising the sequence $CX_6C)$. In other embodiments, the 25 molecule is a peptide analog or derivative. In a preferred embodiment, the substrate is a solid phase on which the molecule (e.g., peptide) has been synthesized, with a high degree of purity. In specific embodiments, the REDOR measurements of the molecule on the substrate can be done in 30 a dry nitrogen atmosphere, under hydrated conditions, and when the molecule is either free or bound to a target. invention is also directed to a solid phase substrate having a surface to which is attached a population of molecules (preferably peptides, peptide derivatives, or peptide 35 analogs), suitable for obtaining REDOR NMR measurements of the molecules. In specific embodiments, at least 90% of the population consists of a single molecule (i.e., 90% purity).

```
In a more preferred aspect, solid phase substrates, as described above, of producing such solid ed.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Methods
                                              In a more preferred aspect, and the entire e
                                                                                                                            also provided. spectroscopy is performed on the same spectroscopy is p
                                                                  Step 46 REDOR spectroscopy is performed on the binder performed sample.

Step 46 REDOR spectroscopy binder prenaration. spectromet binder prenaration. spectromet binder prenaration.
                                                                                                strategically labeled, binder preparation, spectrometer sample preparation nulse semience and excitation nulse details include final and excitation nulse semience
WO 96130849
                                                                                                        details include tuning, and excitation by standard methods out by standard methods parameters and be carried out by standard methods.
                                                                                                                  Parameters and turning, and excitation pulse sequence.

Parameters and turning, and excitation pulse sequence.

Parameters and turning, and excitation out by standard methods.

Preparation can be carried eamning of dried in N. and preparation can be carried eamning.
                                                                                                       binder percide-substrate sample is sealed in he how to limit the percide-substrate sample is sealed in necessary.

The rotor can be cooled.
                                                                                                                                                                                                                    er motion.

An alternative final sample preparation step is to bind:

An alternative to the binder pentide resin sample and
                                                                                                                                                                              An alternative final sample preparation step is to binder nentide optionally. The binder nentide the complex in N..
                                                                                                                                                                                                  then dry the from the highly accurate REDOR NMR distances this alternative.
                                                                                                                                                                                                            can be split from the highly accurate any conformational this alternative, and thus reflect any conformational this bound binder and the first bound because the first boun
                                                                                                                                                                                                                      this alternative, and thus reflect any conformational this alternative, and thus reflect with the target of the bound binder when hinding with the target of the bound of the rear occur.
                                                                                                                                                                                                                                                                                          Jes that resonance, REDOR measurements.

A triple adaptable to REDOR mine is adaptable to nine is adaptable to nin
                                                                                                                                                               binder motion.
                                                                                                                                                            then dry the complex in N2.
                                                                                                                                                                                                                                of the bound binder and thus reflect any contormal thus reflect any target.

Changes .....
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             Such machines
                                                                                                                                                                                                                                                                machine is adaptable to REDOR measurements. Such machine are commercially available collins. Collins. Collins.
                                                                                                                                                                                                                                                                        are commercially available from Bruker (Billerica, Alto, and varian the laborator of the commagnetics (Fort collins, for use is in the laborator of the commagnetics are chemmagnetics.
                                                                                                                                                                                                                                                                                Chemmagnetics (Fort collins to use is in the laboratory of use is in the laboratory of Ithaca, My).

Chemmagnetics (Fort suitable for use My).

An exemplary machine suitable (Ithaca, My).

5 Prof. Zax.
                                                                                                                                                                                                                                                    A triple resonance magic angle spinning machine is adaptable to REDOR measurements.
                                                                                                                                                                                                                                                                                                    Prof. Zax. 7.05 and RF pulse excitation and receiving includes a 7.05 and RF pulse excitation and receiving united kingdom)
                                                                                                                                                                                                                                                                An exemplary machine sultable tor use is in the sultable tors.
                                                                                                                                                                                                                                                                                                              includes a 7.05 Telsa magnet from Oxford Instruments (Caroling and receiving and RF pulse excitation and evamniary and warr are conventional in the NMR are conventional i
                                                                                                                                                                                                                                                                                                                        United Kingdom) and RF pulse excitation and receiving An exemplary the NMR art. MAG nrohe hardware conventional in the resonance. MAG nrohe hardware rorn's a rrinle resonance.
                                                                                                                                                                                                                                                                                                                                  nardware conventional in the NMR art. MAS Probe from measurement rotor is a triple resonance, was probe from chemmaonerics
                                                                                                                                                                                                                                                                                                                                                                                                        nagnetics.

The exemplary magnetic corresponding Larmor fremeroises

The exemplary with the corresponding to the c
                                                                                                                                                                                                                                                                                                                                                                    The exemplary magnetic field is adjusted for a 'H Larmor frequencies with, corresponding Larmor frequency of 300 Mhz with, and 30.4 Mhz, respectively.

Frequency of 75.4 and 30.4 Mhz, and 30.4 Mhz, respectively.
                                                                                                                                                                                                                                                                                                                                                                            frequency of 300 Mhz with, corresponding Larmor frequent with and 30.4 Mhz, respectively. with and 30.4 Mhz, is 4.8 kHz. with frequency (a) is 4.8 kHz.
                                                                                                                                                                                                                                                                                                                                                                                      for "C and "N of 75.4 and 30.4 Mhz, is 4.8 kHz, is respectively. With is 4.8 kHz, is respectively. It is a second respectively
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            15N resonances
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            rotor period (Tr) of 0.208 maec.

The low natural abundance of Alternations

The low ratural corrections
                                                                                                                                                                                                                                                                                                                                                                                                                  are measured background corrections.

The low natural abundance of Min eliminates the Alternatively.

The low natural background corrections.

Alternatively.
                                                                                                                                                                                                                                                                                                                                                                    exemplary probe spin frequency (w, is 4.8 kHz.)

exemplary probe spin period (Tr.) of 0.208 meec.

The period (Tr.) aminasnes of 18.

The period (Tr.) aminasnes of 18.
                                                                                                                                                                                                                                                                                                                       30 Chemmagnetics.
                                                                                                                                                                                                                                                                                                                                                                                                                 are measured.
```

```
measurements can be done with conventional background
                                                                                                                                ections.

REDOR is a pulse NMR technique requiring careful in and is a pulse NMR technique requiring careful in and is a constant in the constant in the constant is a pulse NMR technique requiring careful in and is a constant in the const
                                                                       excitation of appropriate with the MAS rotor Many alternative REPORR

synchronous with the decay.

synchronous induction decay.
                                                                                                                       the are found in the references cited hereinahove excitation sequences found in the references cited hereinahove
                                                                                                                                excitation sequences found in the multiple of which are can involve multiple of the sequences can involve multiple of the references.
WO 96130849
                                                                                                                                         some of which are found in the reference preferred for use in this these sequences can involve multiple pulse sequence preferred for use in the referred for use in this the reference preferred for use in this involve multiple preferred for use in this the sequence preferred for use in this the sequence preferred for use in this this involve multiple preferred for use in this involve multiple preferred for use in this this involve multiple preferred for use in this involve multiple preferred for use involve multiple multiple multiple preferred for use involve multiple mult
                                                                                                              the 12M tree induction decay.
                                                                                                                                                                 Period. The simple pulse sequence for 8 rotor periode is invention requires sequence for 8 rotor periode is invention exemplary sequence.
                                                                                                                                                                                       The exemplary sequence for 8 rotor periods is manner an an is detailed herein program an an is detailed herein program an is detailed in the NMR arts can program an illustrated in skilled in the NMR arts can be skilled in the such that
                                                                                                                                                                                                illustrated in Fig. 4, and is detailed herein in a manner rhree channels excitation that those skilled in measurement. Three channels excitation that the such that the for similar measurement.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       Three channels excited
                                                                                                                                                                                                                      spectrometer for similar measurement. S1, and to the resonances the inc channel supplies are runed to the resonance supplies are runed to the resonances.
                                                                                                                                                                                                                               are the "H channel whose distance is to be measured."

The nuclei whose distance is to be measured.
                                                                                                                                                                                                                                            The incled whose is initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the resonance of the nuclei whose initially tuned to the nuc
                                                                                                                                                                            such that those skilled in hearned to similar measurement.

15 spectrometer for similar head in channel
                                                                                                                                                                                                                                                       of the nuclei whose distance is to be measured. The time sequence.

The time sequence to be measured. The time sequence to be measured.

The interest time is to be measured.

The interest to be measured.

The interest to be measured.

The interest time is to be measured.

The time sequence.
                                                                                                                                                                                                                                                                               Proton coupled to the right) of the exciting illustrated.

Proton coupled to the right) of the exciting illustrated.

Proton coupled to the right) of these channels is illustrated.
                                                                                                                                                                                                                                                                                        (increasing to the right) of the exciting signals (inc. the right) of the exciting illustrated.

(increasing to the right) of these channels is illustrated.

(increasing to the right) of these channels is illustrated.

(increasing to the right) of the excitant on is and initial excitation is and initial excitation.
                                                                                                                                                                                                                                                                                                                                                          ically) in each of these channels is illustrated. It is applied to is applied to initial with an initial and initial excitation is applied to initial excitation is applied to initial with manners: either an initial with manners: either an initial with manners: either an initial in the in either of two manners:
                                                                                                                                                                                                                                                                                                            In the in either of two manners: either and preferred, as illustrated and preferred as pulse may be applied or.
                                                                                                                                                                                                                                     channel RF Power 15 1nltlally tuned to
the short of interest.

20 Proton coupled to the short of the short of
                                                                                                                                                                                                                                                                                        the way be applied transfer from the protons is made.

25 pulse polarization transfer from the protons polarization transfer from the p
                                                                                                                                                                                                                                                                                                                                               cross polarization transfer from the protons is made. the 'H

cross polarization transfer applied at time achieve a

cross polarization transity is applied to achieve a

sufficient RF intensity is respectively. to achieve a

sufficient RF intensity is respectively.
                                                                                                                                                                                                                                                                                                                                    pulse may be applied or as illustrated and preferred.

pulse may be applied transfer from the protons is made.

cross polarization transity is applied at time 54 in both
                                                                                                                                                                                                                                                                                                                                                          Sufficient RF intensity is applied at time 54 in both to achieve a flip time of so and 51 respectively! flip time of and the channels!
                                                                                                                                                                                                                                                                                                                                                                   and what channels to the initial what excitation, subsequent to the initial usec.
                                                                                                                                                                                                                                                                                                                                             Hartman-Hahn precession match at a 7 spin flip time of 13.2

Hartman-Hahn precession to the initial with the MAS probe rotor for N

Hartman-Hahn precession bhase with the MAS probe rotor for N

Dulses 56 are applied in phase with the MAS probe rotor for N
                                                                                                                                                                                                                                                                                                                                                                                        Heec. Subsequent to the initial with the MAS probe rotor for No pulses of are applied by line 59. With sufficient RF pulses cycles.
                                                                                                                                                                                                                                                                                                                                                                                                      Pulses 56 are applied in phase with the MAS Probe RE

Pulses 56 are denoted by line flip time of 13 2 "esc

Totor cycles, achieve a m spin flip time of intensity to achieve a mean flip time of intensity time of 
                                                                                                                                                                                                                                                                                                                                                                                                                 rotor cycles, denoted by line spin flip time of 13.2 µBec.

Totor cycles, achieve a m spin flip evstematically to reintensity to achieve a pulses is varied evstematically to achieve a pulses is partied evstematically to achieve a pulses is varied evstematically to achieve a pulse evstematically achieve achieve achieve a pulse evstematically achieve a pulse e
                                                                                                                                                                                                                                                                                                                                                                                                                          intensity to achieve a T spin flip time of 13.2 peec reduce to achieve a T spin flip time of these T pulses is waried systematically to the nmm are well known in the nmm are phase of these manner well known in the nmm are spin a manner well known in the nmm are spin a manner well known in the nmm are spin a manner well known in the nmm are spin a manner well known in the nmm are spin a manner well known in the nmm are spin a manner well known in the nmm are spin a manner well known in the nmm are spin a manner well known in the nmm are spin a manner well known in the nmm are spin a manner well known in the nmm are spin a manner well known in the nmm are spin a manner well known in the nmm are spin and the nmm are 
                                                                                                                                                                                                                                                                                                                                                                                                      phase of these well known in the NMR arts.

35 artifacts in a manner well known in manner are derailed in manner.
                                                                                                                                                                                                                                                                                                                                                                                                                                                  archiacts in a manner well known in Table 1.
```

Table 1

	¹⁵ N π Pulse Ph	ase Sequencing
5	Number of rotor cycles between excitation and observation	Phase sequence (in processing frame)
	2	YY
	4	XYXY
	8	XYXYYXYX

10

The phase sequence is expressed as the axis, in the frame processing with the ^{15}N spins, about which the π spin flip is made. This axis is systematically varied depending on the number of rotor periods intervening between the ^{15}N excitation and signal observation. The illustrated phase sequences may be varied into equivalent sequences in a conventional manner. For example, "XYXY" is equivalent to "-YX-YX". Finally, at 501 the free induction decay of the ^{15}N spins is observed and generates the time domain output signal.

In the ^1H channel, the preferred sequence is an initial exciting $\pi/2$ pulse 53 followed with the previously described cross polarization transfer 54 to the ^{15}N spins. The less preferred sequence omits these initial pulses in favor of a $\pi/2$ ^{15}N excitation. During the subsequent spin evolution time 25 for N_c rotor cycles and the free induction decay time 501, a decoupling field 55 is applied to the protons. The preferred decoupling field has a 66 kHz RF intensity to achieve a ^{1}H π spin flip in 7.6 μ sec.

In the 13 C channel, two distinct options must be 30 measured. The first option (not illustrated) has no 13 C exciting pulses. The second option (illustrated) has synchronous π pulses 57 applied for N_c rotor cycles at the rotor frequency but with a fixed phase delay 58, denoted by t_1 , and at sufficient signal intensity sufficient to achieve a 35 π spin flip time of 10.6 μ sec. Any value of t_1 may be used; the preferred value is 1/2 the rotor period, $T_c/2$.

Alternative REDOR pulse sequences include 2 or more "C pulses per rotor cycle.

Summarizing still with reference to Fig. 4, a REDOR measurement scan is characterized by the number of rotor 5 cycles, N_c , of spin evolution. A complete scan comprises, first, an equilibration period, preceding the illustrated pulse sequences. Second, there is a ^{15}N excitation period comprising pulses 53 and 54. Third, there is a spin evolution period for N_c rotor cycles which has two options,

- 10 both measured. Both options comprise the application of decoupling 1 H field 55 and synchronous in phase 15 N π pulses 56. The first option has no 13 C excitation; the second has synchronous phase displaced 13 C π pulses 57. Fourth, and finally, there is observation of free induction decay 501 of
- 15 the ^{15}N spins. Fig. 4 illustrates an N_c of 8. Each scan option is repeated, and the induction decay signal accumulated, for a sufficient number of times to obtain acceptable signal to noise ratio. With the preferred practice, this has required less than approximately 5,000 scans, and typically 3000 have been sufficient.

An alternative implementation of the REDOR measurement interchanges the roles of ¹³C and ¹⁵N and measures the free induction decay of ¹³C. Further, the invention is not limited to this described pulse sequence and is adaptable to

25 equivalent pulse sequences yielding direct inter-nuclear dipole-dipole interaction strengths.

Following REDOR measurement step 46, is data analysis step 47. This comprises several substeps. As is conventional, the free induction decay signal is Fourier transformed from the time domain to the frequency domain. The scan option without the ¹³C excitation produces a transformed signal with an observed ¹⁵N resonance peak of magnitude S; the scan option with ¹³C excitation produces an observed ¹⁵N resonance peak of magnitude S_f. The REDOR output signal, denoted ΔS/S, is conventionally formed according to the equation:

$$\frac{\Delta S}{S} = \frac{(S - S_f)}{S} \tag{2}$$

The output signal is observed for different N_c . Preferably 0, 5 2, 4, and 8 rotor cycles are observed. Other preferred N_c will be apparent during the following description.

Further analysis of the REDOR output signal, ΔS/S, is made clearer by a very brief explanation of how this output signal represents the spin 1/2 dipole-dipole interaction between the ¹³C and ¹⁵N. In the spin evolution period, the ¹H decoupling excitation eliminates all proton effects from the ¹³C and ¹⁵N NMR spectra. Magic angle spinning, in the scan option without any ¹³C excitation, eliminates all nuclear dipole-dipole and chemical shift anisotropy from the NMR line. Thus signal S represents an NMR resonance without any dipole interaction. However, in the second scan option, the ¹³C π spin flip pulses reintroduce in a controlled manner the dipole-dipole interaction. This interaction causes additional dephasing, or loss of signal strength, in the observed ¹⁵N signal. Thus signal S, represents an NMR

additional dephasing, or loss of signal strength, in the

20 observed ¹⁵N signal. Thus signal S, represents an NMR
resonance with dipole interaction and the output signal ΔS/S
represents the percentage strength of pure dipole-dipole
interaction between the ¹³C and ¹⁵N nuclei. The exact loss of
signal strength depends on the timing of the ¹³C pulses and
the number of rotor cycles for which they are applied.

In the alternative where a general phase delay, t₁, is used, the expression for the REDOR signal is derived by numerically integrating the following equations from the Pan et al. reference (1990, J. Magnetic Resonance 90:330-340):

$$S_{f} = 1 - \frac{1}{2\pi} \int_{0}^{\frac{\pi}{2}2\pi} \cos \left[T_{r} \omega_{D}^{\prime}(\alpha, \beta, t_{1}) \right] \sin \beta d\beta d\alpha$$
 (3)

35 where

30

$$\omega_{D}(\alpha,\beta,t) = \pm \frac{1}{2} D_{CN} \{ \sin^{2}(\beta) \cos 2(\alpha + \omega_{r}t) - \sqrt{2} \sin 2\beta \cos(\alpha + \omega_{r}t) \}$$

$$\omega_{D}'(\alpha,\beta,t_{1}) = \frac{1}{T_{r}} \{ \int_{0}^{t_{1}} \omega_{D}(\alpha,\beta,t') dt' - \int_{t_{1}}^{T_{r}} \omega_{D}(\alpha,\beta,t') dt' \}$$

$$(4)$$

This integration can be done by standard numerical integration techniques such as are found in Press et al.,

5

25

- Numerical recipes: the art of scientific computing,
 Cambridge, U.K., Cambridge University Press, (1986), chapter
 4, which is herein incorporated by reference. Alternatively
 the expression can be directly evaluated from the symbolic
 representations by numerical tools such as Mathematica from
- Mathsoft Inc. (Champaign, IL) or Mathcad from Mathsoft Inc. (Cambridge, MA). In a preferred embodiment, however, a much simpler approach is used.

In the preferred embodiment, the ^{13}C pulse phase delay is 1/2 the rotor period, T_r , and the preceding equations can be simply expressed (Mueller et al., 1995, J. Magnetic Resonance, in press):

$$\frac{\Delta S}{S} = 1 - [J_0(\sqrt{2})\lambda]^2 + 2\sum_{k=1}^{\infty} \frac{1}{16k^2 - 1} [J_k(\sqrt{2}\lambda)]^2$$

$$\lambda = N_c T_r D_{CN}$$
(5)

where J_k is a Bessel function of the first kind. Adequate accuracy is obtained by limiting the summation of equation 5 to its first five terms. Fig. 5 is a graph of this equation. Vertical axis 61 represents $\Delta S/S$; horizontal axis 62 represents λ ; and graph 63 represents equation 5.

In detail, step 47 of Fig. 3 uses equation 5 and the REDOR output signal, $\Delta S/S$, for various values of N_c to obtain a best value for D_{CN} , the dipole interaction strength. The internuclear distance is simply and directly determined from D_{CN} by equation 1. An exemplary method for finding the best

value of D_{CN} is to use a least squares method. First, form the sum of the squares of the differences of the observed $\Delta S/S$ and $\Delta S/S$ computed from equation 5, which will be a function of D_{CN} , T_r , and N_c through λ . Second, find the value 5 D_{CN} minimizing this function by searching exhaustively in sufficiently small increments over the relevant range. For example, D_{CN} can be varied by varying R in 0.01 Å increments from 0.5 to 8 Å. More efficient minimization methods as presented in Press et al. chapter 10 can also be used.

- 10 Values of the Bessel functions can be simply calculated by the methods in Press et al, supra, § 6.4. Alternatively, this minimization and best value determination is easily performed directly from the symbolic representations with the previously cited mathematical packages.
- The example in Section 6.6 provides typical results of this measurement and analysis method.

This completes the method of Fig. 3 and determines the internuclear distance between the ¹³C and ¹⁵N nuclei to which the excitation channels were tuned for the REDOR NMR

20 measurements. If other C-N pair distances are to be determined in the labeled binder, step 46 as detailed above is repeated for the other distinct resonances. If the alternative ¹⁵N resonances cannot be distinguished, separately labeled binders are prepared and measured.

25

5.7. CONSENSUS, CONFIGURATIONAL BIAS MONTE CARLO Broad overview

With reference to Fig. 1, having found N specifically binding members of one or more libraries, step 2, selected a 30 candidate pharmacophore shared by all these binders, step 3, and determined a few strategic distances in the vicinity of the candidate pharmacophore, step 4, precise pharmacophore and binder peptide structures are now determined by the preferred method, the consensus, configurational bias Monte 35 Carlo method. Other orderings and identities of these steps are possible. For example, the binders may be predetermined thereby rendering step 2 unnecessary. Further, no strategic

distance measurements may need to be made, and step 4 may be omitted. Alternatively, a partial structure determination step may be inserted before step 4 to guide selection of distances for measurement.

- Pharmacophore structure determination of this invention is not limited to the CCBMC method to be described. CCMBC makes the most efficient use of heuristic consensus binding and partial distance measurement information. However, the consensus pharmacophore can be determined by methods
- 10 including but not limited to use of exhaustive REDOR NMR measurements or by extensive but fewer REDOR measurements in conjunction with a conventional molecular structure determination method, such as molecular dynamics, conventional Monte Carlo, or even peptide folding rules.
- In the following description, the CCBMC method is broadly overviewed; subsequently, details of important steps are described; and finally a description of the preferred computer method and apparatus for practicing the invention is given. From the description of the methods, equations, data
- 20 structures, and programs provided herein, one will be able readily to translate them into implementations.

Although the following descriptions are directed to binders isolated from the preferred library of peptides comprising the sequence CX₆C (constrained by disulfide bonds),

- 25 the method is applicable to more general organic diversity library members. It is immediately applicable to compounds from constrained peptide libraries with other scaffolds and also to compounds from similar peptoid libraries. It will be readily apparent that the method is applicable to any
- 30 compounds whose structural region of interest exhibits conformational degrees of freedom at a temperature of interest (e.g., body temperature -- 37°C) that are limited to torsional rotations of rigid molecular subunits about bonds between the subunits, in which any loops present in the
- 35 structural region of interest are independently rotatable by concerted rotation (see Section 7. Appendix: Concerted Rotation). Examples of such compounds include but are not

limited to peptides, peptoids, peptide derivatives, peptide analogs, etc., including members of libraries discussed in Section 5.2, supra.

General features of Monte Carlo simulation methods are known. A reference is Rowley, Statistical mechanics for thermophysical property calculations, Englewood Cliffs, N.J., PTR Prentice Hall (1994), especially chapters 5 and 7, which is herein incorporated by reference. The application of simple Monte Carlo to constrained peptides has conventionally been hindered by difficulty generating geometrically proper and energetically useful conformational alterations, and by the consequent wasteful and inefficient exploration of conformational space. This method overcomes these problems for constrained peptides with a novel combination of techniques. In addition, this method is uniquely able to incorporate partial information about binding affinities and distance measurements to improve determination of the pharmacophore structure, one goal of the invention.

Fig. 8 is a overview of the method. Step 91 represents
20 the initial geometric and chemical structure of each binding
peptide in computer memory. Peptide geometric structure is
represented as a set of records, each record representing one
rigid subunit or one atom of the peptide. The subunit
records are linked together as the subunits are linked in the
25 peptide molecule. Each rigid unit record includes fields for
the composition, structure, and connectivity of the rigid
unit represented. Since the rigid units only undergo
torsional rotations about mutual bonds, their internal
geometric structure is fixed.

If a previous run with these peptides has been done, peptide initial structure may be chosen as one of the structures generated late in that run. Such an initial structure is desirable since the effects of arbitrary initial conditions have been eliminated. Alternatively, an initial structure is generated from a prototypical backbone without side chains by adding sidechains with random torsional orientations. For members of each type of diversity library,

a prototypical backbone meeting structural constraints and representing an allowed configuration for a member possessing no side chains can be defined. The prototypical backbone for the CX,C library is generated from the CCBMC model itself as 5 run for the linear peptide C(gly) 6C (SEQ ID NO:7) using a Hamiltonian consisting only on the H_{NPR} term. The H_{NPR} term contains only terms which, in the disulfide bond backbone region $-C_1-S_1-S_2-C_2-$, limit the S_1-S_2 distance to 2.038 Å and both the C_1 - S_2 and the S_1 - C_2 distances to 2.883 Å. When run 10 for a linear peptide, no Type II backbone moves are made. Only Type I backbone moves which remove and regrow randomly selected portions of the backbone are used to generate backbone alterations. The model is run with temperatures gradually decreasing from room temperature to a small 15 temperature, approximately 1 °K. The final low temperature structure is used for the prototyptical backbone. Backbones for similar constrained peptide libraries can be constructed in similar manners.

In memory, for each peptide, a current structure is

20 represented; the initial current structures being the just
assigned initial structures. Also in memory is represented a
proposed modified structure for one peptide. At step 92 the
processor generates "moves" that transform the current
structure of a randomly chosen peptide into a proposed

25 modified structure. The moves mimic body temperature (37 °C)
thermal agitation experienced by the binders so that their
equilibrium structure may be determined.

Generation of these moves for conformationally constrained peptides is an important aspect of this method.

30 There are two move types. Type I moves alter the conformation of the side chain of a randomly chosen amino acid of the randomly chosen peptide. The alteration is built by side chain removal followed by side chain regrowth into a new torsional conformation. During regrowth, unfavorable overlap with neighboring side chains is avoided. Type II moves alter the conformation of a limited random region of the peptide backbone of a randomly chosen binder by

performing linked, or "concerted", rotations, the linking being such that only four backbone rigid units are spatially displaced. Thereby the internally bonded ring of 8 amino acids will not be disrupted. A reference describing a 5 similar move in linear alkane molecules is Dodd et al., A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses, Molecular Phys., vol 78, pp 961 et seq. (1991), which is herein incorporated by reference. The ratio between the Types I and II moves is an 10 adjustable parameter with a preferred value of 4.

Another important aspect of this method is that both moves are selected in a "configurationally biased" manner. Normal Monte Carlo methods use standard Metropolis procedures, in which each proposed structure is generated 15 randomly and independently of the current structure with an equal a priori probability. However, for complex molecules, it is known that this typically results in the generation of many highly improbable or energetically unlikely structures. In some situations up to 105 wasted moves are generated for 20 each useful move, a very considerable waste of processor resources. In contrast, the method of this invention generates proposed structures according to an a priori probability depending on the current structure and the energetic cost of the new structure. This bias toward more 25 acceptable structures of lower energy avoids generating highly improbable structures, making a very much more efficient use of processor resources. Because detailed balance must be satisfied, the acceptance probability of the configurationally biased method must include factors in 30 addition to the usual Boltzman factor. A reference applying a similar method for simple linear alkanes is Smit et al., Computer simulations of the energetics and siting of nalkanes in zeolites, J. Phys. Chem. vol 98, pp 8442 et seq. (1994), which is herein incorporated by reference.

At Step 93 the processor evaluates the energy, or Kamiltonian, of the proposed configuration. The Hamiltonian contains two groups of terms: conventional physical energy

terms, and heuristic constraint terms. Conventional terms include the energies of rigid unit torsional rotations and of Lenard-Jones, electrostatic interactions, and H-bonding between atoms in different rigid units. Bond lengths and 5 angles are assumed fixed at the temperature of interest and their energies constant. These conventional interactions are exclusively intramolecular; no physical intermolecular interaction effects are considered in this invention. References for the conventional energies are Weiner et al., An all atom force field for simulations of proteins and nucleic acids, J. of Computational Chem., 7:230-52 (1986); and Weiner et al., A new force field for molecular simulation of nucleic acids and proteins, J. Amer. Chem. Soc. 106:765 (1984) (herein referred to as the "AMBER references"), which

15 are herein incorporated by reference. Another important aspect of the Monte Carlo method of this invention is the heuristic terms: the consensus term and the measurement constraint term. They uniquely make use of partial information on the binder peptides to guide the Monte 20 Carlo simulation. The consensus term, H_{consensus}, is added to the Hamiltonian to represent that all the binders do in fact bind to the same protein target in the same physical and chemical manner. Since binding occurs at the shared candidate pharmacophore in each binder, this term makes 25 energetically unfavorable moves that cause the geometric structure in the shared pharmacophore to depart from an average, common structure. Pseudo chemical "bonds" to this average structure are added which mimic the actual physical bonding to the surface groups of the protein target. 30 candidate pharmacophore is in fact the actual pharmacophore, this energy will become minimized and small in the equilibrium configuration, since there will be an actual, shared, geometric configuration. If the candidate pharmacophore is not the actual one, this term will not 35 become minimized or small, as there is no physical reason for this region of the peptide molecules to share a common structure. This is the only Hamiltonian term which couples

the N binders together; no physical intermolecular effects are considered. The binders are otherwise treated independently by the method.

The measurement constraint term, Hnng, is added to 5 represent the distance measurements made, which are in fact actual distances in the molecules and constrain any simulated structure. This term makes energetically unfavorable, by adding pseudo chemical bonds of the measured lengths, moves that cause the constrained internuclear distance to depart 10 from their measured values. Of course if no partial distance measurements have been made or are otherwise available, this term may simply be omitted from the Hamiltonian without adversely affecting the practice of this step. Which measurements to make, if any, is guided by the results of the 15 consensus structure determined. If an adequate structure can be obtained without assistance of distance measurements, none need be incorporated. If inadequate results are obtained, additional iterations of the method will need distance measurement inputs.

Step 94 tests the proposed structure against an acceptance probability, accept(curr->prop). This acceptance probability is determined by the energy of the proposed structure previously computed in step 93. If the proposed structure fails this test and is not accepted, the method progresses immediately to step 96. If the proposed structure meets the test and is accepted, the accepted proposed structure replaces and becomes the current structure. The proposed structure of this peptide is also saved (given certain other conditions detailed later) in a separate memory store of structures for later analysis. This structure store is preferably on disk.

Repeated application of the concerted rotation may lead to a slightly imperfect structure, due to numerical precision errors. In an alternative embodiment, peptide geometry would 35 be restored to an ideal state by application of the Random Tweek algorithm after several thousand moves (Shenkin et al., 1987, Biopolymers 26:2053-85).

Step 96 tests whether enough structures of equilibrated total energy have been generated in this simulation run. The run terminates if a sufficient number have been generated. Sufficiency is determined on the basis of whether the statistical sampling errors of the average pharmacophore structure determined at step 97 is adequate (typically, less than 0.25 Å). Preferably, 25,000 equilibrated structures would be accumulated for each run. Also, preferably, three runs would be performed for a total of 75,000 saved

Fig. 9 illustrates energy equilibration of an actual run. Axis 101 is the total energy of a set of peptide binders; axis 102 is the number of moves accepted. Traces 103 represent total energies of all binders from each of the 15 three runs. Typically, run energy rapidly equilibrates within less than approximately 2000 moves in most cases. Subsequent saved structures are counted toward termination. Traces 103 display typical energy variations superimposed on a secular stability. The illustrated energy variations 20 typically comprise several components having different variabilities. First, there is a very high frequency oscillation with a period of a few tens of moves (known as "hair"). Second, there is a low frequency oscillation with a period of several hundred to a few thousand moves and with 25 low amplitude.

Step 97 analyzes the structure stored in memory. In the simplest preferred embodiment, the stored geometric structures for each binder are simply averaged, yielding a final structure for each binder and for the candidate

30 pharmacophore. In another alternative, clustering software seeks clusters of similar structures for each binder. The clusters are then averaged to give a final structure for each variant structure for each binder. The variants represent alternative foldings for the binder. Exemplary clustering

35 methods are found in Gordon et al. Fuzzy cluster analysis of molecular dynamics trajectories, Proteins: Structure, Function and Genetics 14:249-264 (1992).

Alternative post-processing can be done on the clustered structures to account for small bond angle vibrations. vibrations are expected to make small perturbations to the clustered structures determined by the Monte Carlo method and 5 can be accounted for by a brief molecular dynamics simulation. Such a simulation is fully defined by the Hamiltonian, comprising the physical and heuristic energies to be described infra in Eqn. 8, and by the temperature of interest. The structures observed during the simulation are 10 averaged to determine a final more accurate equilibrium structure. A code capable of performing such a simulation is Discover® from BIOSYM (San Diego, CA). Preferably, the molecular dynamics simulation would be run for approximately 105 bond angle vibration periods. Since the typical bond 15 angle vibration period is 10^{-2} ps (1 ps = 10^{-12} sec.), such a run will encompass approximately 1 ns of molecular time.

Configurational bias move generation details

- One Type I or II move will, in general, alter the 20 position of several rigid units on a side chain or along the backbone. Each altered rigid unit is sequentially considered during move generation. The Hamiltonian describing the energy of the rigid unit currently being considered in a move is divided into an internal, uint, and an external, uext, part,
- 25 where uext is all energy not included in uint. In the preferred embodiment, uint is set to 0; an alternative choice would be to include only the torsional interaction energy between this rigid unit and units to which it is currently bound. uint generates a probability distribution, pint, according to which
- 30 is generated a set, ϕ_k , k = 1...K, of candidate torsional angles for the bond between the rigid unit being examined and rigid units already examined. u^{ext} generates another probability distribution, p^{ext} , according to which is selected one torsional angle from the prior set as the proposed new
- 35 angle for the rigid unit being examined. These probabilities are defined by the equations:

5

$$p_{i}^{int}(\phi_{i,k}) \propto \exp\left[-\beta u_{i}^{int}(\phi_{i,k})\right]$$

$$p_{i}^{ext}(\phi_{i,k}) = \frac{\exp\left[-\beta u_{i}^{ext}(\phi_{i,k})\right]}{w_{i}^{ext}}$$

$$w_{i}^{ext} = \sum_{k=1}^{K} p_{i}^{ext}(\phi_{i,k})$$
(6)

In this equation, "," signifies the rigid unit being considered, K is the total number of candidate torsional angles generated by p^{int}, and β = 1/kT (k is Boltzman's constant; T the temperature, preferably 37 °C). The overall probability of generating a transition from the current to the proposed structures and accepting the proposed structure are given by the equations:

$$P(curr-prop) \propto \prod_{i=1}^{K} p_i^{int}(\phi_{i,k}) p_i^{ext}(\phi_{i,k})$$

$$W^{new} = \prod_{i=1}^{K} w_i^{ext}$$

$$accept(curr-prop) = min(1, \frac{W^{new}}{W^{old}})$$
(7)

25

30

In this equation, M is the total number of rigid units added in the move. W^{old} is a weight for the reverse move and will be described subsequently.

Because energy is included in the generation probabilities, proposed structures are preferentially of lower energy. Since the acceptance of proposed structures depends on their energies, the acceptance of proposed structures is thereby more probable.

Peptide memory representation details

It is well known that at body temperature peptides consist of linked rigid units capable only of torsional rotational about mutual bonds whose lengths and angles are 5 fixed. The torsional rotations respect any molecular conformational constraints. See Cantor et al., Biophysical chemistry part I the conformation of biological macromolecules, New York, W.H. Freeman and Co. (1980), which is herein incorporated by reference. Table 2 lists the rigid 10 units encountered in the preferred embodiment of this invention utilizing libraries of conformationally constrained peptides. Table 2, where applicable, also lists dihedral bond angles between incoming and outgoing bonds to a rigid unit and the assigned unit type.

15

Table 2

Chemical Bond angle Type (if applicable) Structure Backbone and side chain 20 rigid units -NH, Α 70.5° В -CaH-70.5° - CONH -C 25 D -COOH Side chain only rigid units 70.5° -CH,-E 70.5° F 30 -CH-70.5° **-**S-G 00 -C₆H₄-Η -CH, I -OH J K -SH

Туре	Chemical Structure	Bond angle (if applicable)
L	-NH ₂	
М	-C ₆ H ₅	
N	- CONH ₂	
0	-CN3H4	
P	-C,N2H,	
Q	-C,NH6	

Table 3 illustrates the decomposition of all amino acid side chains into rigid units. Glycine is a special case, without a side chain. Proline is a special case with a side chain cyclically bonded to the backbone amino N.

Table 3

	Amino Acid	Rigid Units
	Glycine	-CαH ₂ - (SPECIAL CASE)
5	Alanine	-СН,
	Arginine	-CH ₂ -CH ₂ -CH ₂ -CN ₃ H ₄
	Aspartate	-CH ₂ -COOH
	Asparagine	-CH ₂ -CONH ₂
10	Cysteine	-CH,-SH
	Glutamate	-CH ₂ -CH ₂ -COOH
	Histidine	-CH ₂ -C ₃ N ₂ H ₃
	Isoleucine	-CH(-CH ₃)-CH ₂ -CH,
	Leucine	-CH,-CH(-CH ₃),
15	Lysine	-CH ₂ -CH ₂ -CH ₂ -CH ₂ -NH ₂
	Methionine	-CH ₂ -CH ₂ -S-CH ₃
	Phenylalanine	-CH,-C ₆ H ₅
20	Serine	-CH,-OH
	Threonine	-CH(-CH ₃)-OH
	Tryptophan	-CH ₂ -C _e NH ₆
	Valine	-CH(-CH ₃)-CH ₃
25	Tyrosine	-CH ₂ -C ₆ H ₄ -OH

Fig. 10 illustrates a structurally correct but geometrically inaccurate decomposition of the peptide backbone CX₆C into rigid units (inessential hydrogens have been omitted). Rigid units are set off in boxes 121 and their types 122 are indicated. Fig 11 illustrates a structurally correct but geometrically inaccurate decomposition of the peptide backbone and side chains of -arginine-glycine-aspartate- ("RGD") into rigid units. Rigid units are set off in boxes 131 and their types 132 are indicated.

Rigid units are represented as records in memory. The data structure for a peptide comprises records for its constituent rigid units linked together by data pointers exactly as the actual rigid units in the peptide are 5 chemically linked. The record representing a rigid unit comprises fields for: type of the unit, pointers to chemically bonded units, all atoms of the unit and their spatial positions, atoms of the unit that are the target of the incoming and outgoing bonds, amino acid to which the unit 10 belongs, and atomic composition of the unit.

A known, conventional representation of atoms and atomic interactions is taught by the AMBER references. Each atom is divided into a series of subtypes of specific properties. For example, for carbon there are subtypes C, C2, CA, CT, 15 etc.; for nitrogen, there are N, N2, etc.; for oxygen, there are O, O2, etc.; and for hydrogen, there are H, H2, etc. Bonds between each pair of subtypes are separately characterized by equilibrium lengths, angles, and torsional energies. Interactions between each pair of subtype atoms 20 are separately characterized by Lenard-Jones force parameters, hydrogen bonding force parameters, and electrostatic charges. Amino acid charge distributions are in Weiner et al., J. of Computational Chem., 7:230-52 (1986).

- Thus each atom in each rigid unit is represented by an in-memory record comprising fields for: its AMBER reference subtype and any electrostatic charge. The atom's spatial position relative to its containing rigid unit, stored in that unit's record, is geometrically determined from the unit's internal chemical structure and bonds by the AMBER bond lengths and angles defined for each of these bonds. The relative spatial positions of atoms within a rigid unit are, of course, fixed, and there is no interaction energy to consider between atoms within a rigid unit.
- Fig. 11 is a complete memory representation of a tripeptide sequence -RGD- (a known pharmacophore). Rigid units are set off in boxes 131 and their types 132 are

indicated. The torsional degrees of freedom between the rigid units are indicated by angle arrows 133. AMBER atoms types are indicated as at 134. Net atomic charges are indicated only for arginine as at 135. Rigid unit records are linked into a data structure modeling the rigid unit's physical linkages. Not shown are relative atomic spatial positions represented by the atoms rectangular coordinates.

All parameters defining the AMBER atomic representations and interatomic forces can be found in Weiner et al., J. of 10 Computational Chem., 7:230-52 (1986), and Weiner et al., J. Amer. Chem. Soc., 106:765 (1984). Conventionally, these parameters are obtained from computer readable files from commercial sources. The preferred computer readable source of these parameters is from Insight II® 2.3.5 software from 15 BIOSYM (San Diego, CA). Other sources are Tripos (St. Louis, MO) and CHARMM (Molecular Simulations, Inc., Burlington, MA).

Interaction energy evaluation details

The form of the intramolecular energy, or Hamiltonian, 20 evaluated at step 93, is an important element of this invention. The Hamiltonian consists of the components:

$$H_{total} = \sum_{l \in binders} H_{l, total}$$

$$H_{l, total} = H_{l, molecular} + H_{l, MMR} + H_{l, consensus}$$
(8)

25

The $H_{l,molecular}$ component is determined from the Weiner et al. references, J. of Computational Chem., 7:230-52 (1986), and J. Amer. Chem. Soc., 106:765 (1984).

30
$$H_{1.molecular} = \sum_{\substack{n: 1 \in \\ rigid unit \\ torsional \\ angles}} \frac{V_{in}}{2} \left(\cos\left(n\phi_{1,i} - \gamma_{i}\right) + 1\right) + \sum_{\substack{i \in j \\ i,j \in \\ atom \ pairs}} \left[\frac{A_{ij}}{R_{1,ij}^{12}} - \frac{B_{ij}}{R_{1,ij}^{6}}\right]$$

$$\sum_{\substack{i \in j \\ i,j \in \\ atom \ pairs}} \left[\frac{q_{i}q_{j}}{\epsilon R_{1,ij}}\right] + \sum_{\substack{i \in j \\ i,j \in \\ atom \ pairs}} \left[\frac{C_{ij}}{R_{1,ij}^{12}} - \frac{D_{ij}}{R_{1,ij}^{10}}\right]$$
35

Here, $\phi_{i,i}$ is the i'th torsional angle between rigid units of the l'th binder peptide, and $R_{i,i}$ is the interatomic distance between the i'th and j'th atoms in different rigid units of the l'th binder. The first term in this equation is the torsional energy of rigid units; the second is the interatomic Lenard-Jones energy; the third is the interatomic electrostatic energy; and the fourth is the interatomic hydrogen bond energy. Rigid unit torsional rotations directly change the first term. Such rotations indirectly thange all other terms as interatomic distances change.

The AMBER parameters V_{in} , A_{ij} , B_{ij} , q_i , C_{ij} and D_{ij} are obtained as stated above. The effect of water is approximated in a known manner by setting ϵ equal to $4\epsilon_0 r$, where r is distance (in Å) in the electrostatic term and ϵ_ϵ is the vacuum permeability.

The distance constraint term, as described, makes energetically unfavorable moves which cause those measured interatomic separations in the simulation to depart from their measured values. If no measured values are available,

- 20 this term is simply omitted from the Hamiltonian. Since this is not a physical energy and in simulation equilibrium the binders should have the measured distance, it is advantageous that this term should make only a small contribution to the equilibrium energy, no more than 10% of the total energy and
- 25 preferably approximately 2.5 to 5%. Further, it is advantageous that the energetic disfavor be weighted by the confidence in the measurements, so that measurements having more confidence have a greater effect.

Many forms of this energy meet these criteria. The 30 preferred form is:

$$H_{1,NMR} = \sum_{\substack{i < j \\ i,j \in \\ observed \\ distance pairs}} \frac{(R_{1,ij} - R_{1,ij}^{(o)})^2}{2w_{1,ij}}$$

$$(10)$$

where R^(o)_{1.13} is a measured distance in the l'th binder peptide between atomic pair ij. This makes the constraints appear as an elastic pseudo-bond with equilibrium length as measured. The w_{1.13} are weights designed to meet the above size criteria.

5 In the preferred embodiment, they are calculated with an overall multiplicative factor limiting the contribution of H_{1.300} to no more than approximately 5% of the total equilibrated energy. Their relative value is selected to reflect the lower reliability of longer measurements. Thus 10 if R^(o)_{1.13} is between 0 and 3 Å, w_{1.13} has a relative value of 1; if the measurement is between 3 and 4.5 Å, the relative value is 2; if between 4.5 and 7 Å, the value is 3; and if the distance exceeds 7 Å, the term is dropped from the sum. Other alternative weight assignments meeting the general 15 criteria are clearly possible.

The consensus constraint term, as described, makes energetically unfavorable moves which cause the candidate pharmacophore in each of the binders to depart from an average, shared configuration. In simulation equilibrium 20 when the candidate is the actual pharmacophore, the binders share the pharmacophore structure and this term should be small. Since this is not a physical energy, in the case where the candidate pharmacophore is correct, this term should not be large compared to the total energy, in equilibrium no more than 10% of the total energy, and preferably approximately 5%. Further, the energetic disfavor should preferably be weighted by the affinity of each binder for the protein target, so that binders with greater affinity have a greater energetic effect.

30 Many forms of this energy meet these criteria. The preferred form is:

$$R_{ij}^{(c)} = \sum_{l \in binders} \frac{R_{l,ij}}{N}$$

$$H_{l,consensus} = \sum_{\substack{i < j \\ i,j \in \\ pharmocophore \\ distance pairs}} \frac{(R_{l,ij} - R_{ij}^{(c)})^2}{2w'_{l,ij}}$$
(11).

R(c),, the shared consensus structure for the candidate pharmacophore, is an average of the interatomic distances between corresponding atomic positions, ij, in the shared pharmacophore in all binders. This makes the constraints 5 appear as a pseudo-bonds to a shared pharmacophore, which represents the binding to the protein target. The w'1.ii are weights designed to meet the above size criteria. preferred embodiment, they are calculated with an overall multiplicative factor limiting the contribution of $H_{1,consensus}$ to 10 no more than approximately 5% of the total equilibrated Their relative value is selected to reflect that binders with lower affinity are less reliable indicators of actual pharmacophore structure. Thus the relative value of the weights is proportional to the logarithm of the affinity 15 of the corresponding binder with an affinity of 1 μ molar having a relative weight of 1. Other weight assignments meeting the general criteria are clearly possible. heuristic H_{consensus} is the only Hamiltonian term linking together the various binders.

20 All Hamiltonian components change only due to the dependence of the interatomic distances, R_{l.ii}, on the rigid unit's torsional rotation. The R_{1,1}, are the well known Euclidean distances between the atomic coordinates stored in the rigid unit records. Calculation of coordinate changes 25 due to rotation of angle ϕ about a bond with unit direction n originating at atom A with position x is well known, but will be detailed. (Throughout, symbols representing vector quantities are indicated by underlining.) First, translate from the current coordinate origin to an origin at position x 30 by adding x to all relevant coordinate vectors. Second, apply a rotation matrix, T, to the atomic coordinate vectors. Third, translate back to the prior coordinate origin from x by subtracting x from all relevant coordinate vectors. A rotation matrix is given by:

$$T = \cos(\varphi) I + nn^{T} [1 - \cos(\varphi)] + M\sin(\varphi)$$

$$\begin{bmatrix} 0 & -n, & n_{\mu} \end{bmatrix}$$

$$M = \begin{bmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{bmatrix}$$
 (12)

5

A reference for this computation is Goldstein, <u>Classical</u>

mechanics, Massachusetts, Addison-Wesley (1981), especially chapter 4, which is herein incorporated by reference.

Type I move generation

Type I moves alter side chain structure of a randomly chosen amino acid in a randomly chosen binder. These random choices are conventionally made by a random number subroutine. The chosen side chain is "removed" from the binder peptide and "grown" back rigid unit by rigid unit. For the next, i'th, rigid unit to be added, K possible new torsional angles are generated according to pint. Preferably 20 K is from 10 to 100. One of these torsional angles is selected according to pext, and the rigid unit is added at this new angle. Determination of pext requires obtaining the normalization w_i^{ext} . At each step the u^{int} and u^{ext} used to calculate the respective probabilities include only interaction energies with rigid units present in other amino acids or already grown back. Rigid units not yet added are ignored. After all the side chain rigid units have been added back, Wnew is computed as the product of the normalization factors.

Fig. 12 illustrates a Type I move for glutamate. At 141 the side chain has been removed. The first $-CH_2$ - unit is added back at 142 with new torsional angle ϕ_1 . The generation according to p^{int} and selection according to p^{ext} of this angle ignores energy interactions with the other side chain rigid units not yet added. At 143, the next $-CH_2$ - rigid unit is added back at angle ϕ_2 . Finally at 144, the last $-CO_2$ rigid

unit is added at angle ϕ_2 . For this last step interaction energies with all the rigid units are considered in generating and selecting the new angle.

Wold is the weight for the reverse move, the move from the proposed new structure to the current configuration. For this, the proposed side chain is removed and regrown in its current structure unit by unit. For the next, i'th, unit generate K-1 possible new torsional angles according to pint, again ignoring interactions with units yet to be added. The 10 K'th new angle is the current angle for that unit. The current torsional angle is selected. Although pext is not used, normalization wiext is determined. After all units have been regrown at the current angles, Wold is computed as the product of the normalizations.

The acceptance probability for the proposed side chain configuration is determined from equation 7 using W^{new} and W^{old}

Type II move generation

Type II moves alter a limited region of the amino acid 20 backbone beginning at a randomly chosen backbone rigid unit of a randomly chosen binder peptide in a manner consistent with conformational constraints due to internal disulfide bonds. These random choices are made similarly to those for Type I moves.

In Type II moves, side chains attached to the altered rigid units move rigidly with their backbone rigid units.

For this move, important geometric constraints must be met. In a randomly chosen binder and at a randomly chosen backbone bond between adjacent rigid units, a torsional angle 30 rotation by ϕ_0 is made. Subsequent backbone torsional rotations are chosen so that a minimum number of rigid units undergo a spatial displacement. This constraint fixes a limited number (if any) of possible subsequent torsional angles as a function of ϕ_0 so that at most 4 rigid units are 35 spatially displaced and rotated with at most 3 additional rigid units undergoing a rotation. This move is an important aspect of this invention and is required to maintain the

PCT/US96/04229 WO 96/30849

conformational constraint due to the disulfide bridge. Since only 7 rigid units are spatially modified, the Type II move preserves the 8 amino acid cycle (20 rigid units), including the cystine side chain.

Fig. 13 illustrates a Type II move of a poly-glycine 7mer. Rigid unit positions are indicated generally by black circles as at 1509 with incoming bonds generally as at 1502. A C rigid unit (B unit) is illustrated in box 1515, and an amide bond (C unit) in box 1516. Backbone structure 1500 in 10 transformed into structure 1501 by the Type II move generated by an initial rotation about bond 1502. Subsequent rotations about bonds 1503, 1504, 1505, 1506, 1507, and 1508 are thereby determined so that the rigid unit 1510 and at most three subsequent units undergo only a rotation without any 15 spatial displacement. The four rigid units between units 1509 and 1510 undergo both a spatial displacement and a rotation as structure 1500 is transformed to structure 1501. No other backbone rigid units are altered.

The derivation of these assertions, including 20 expressions for the allowed angles, is in Section 8. Appendix: Concerted Rotation. Fig. 14 defines notation used in this Appendix: Concerted Rotation. Poly-glycine 7-mer backbone 1600 is the same as in Fig. 13. Rigid unit positions are indicated generally by black circles as at 1601 25 with incoming bonds generally as at 1602. The torsional rotations ϕ_0 to ϕ_6 are about bonds 1602 to 1608, respectively, between sequential, adjacent rigid units. The rigid unit position vectors ro to ro, illustrated as vectors 1610 to 1616, respectively, define the position of these sequential 30 rigid units with respect to a laboratory coordinate system with origin 1609. Summarizing this Appendix, the determination of the fixed torsional angles proceeds as The allowed values for ϕ_1 are the roots of equation 34, which depends on the ϕ_0 driver angle and ϕ_2 through ϕ_4 . 35 But ϕ_2 through ϕ_4 can be determined in terms of ϕ_1 . Two

solutions for ϕ_2 are determined by equation 25 in terms of ϕ_1 . Two solutions for ϕ_1 are determined by equation 29 in terms of

the preceding ϕ 's. Finally, a simple inversion of equation 32 determines one solution for ϕ_i in terms of the preceding ϕ 's. Having found the allowed values of ϕ_1 , then equations 25, 29, and 32 determine corresponding allowed values for the 5 other ϕ 's, which in turn determine the alteration of the first four rigid units caused by the ϕ_0 initial rotation.

More precisely, final torsional angles ϕ_0 to ϕ_6 determine position vectors \underline{r}_i to \underline{r}_i by applying rotation matrix 18 to equations 17 to obtain new position vectors in the laboratory 10 coordinate system, the rotation matrices of equations 16 and 18 being determined by these final torsional angles. Position vectors \underline{r}_c and \underline{r}_s to \underline{r}_7 do not change. Then rigid unit 0 is translated to position \underline{r}_0 ; aligned so that its incoming bond axis is along the direction of the outgoing 15 bond of unit -1; and finally rigidly rotated so that the end of its outgoing bond is at position r. Rigid unit 1 is then translated to position r; aligned so that its incoming bond axis is along the outgoing bond of unit 0; and rigidly rotated so that the end of its outgoing bond is at position 20 \underline{r}_2 . Rigid units 2 to 6 are then added to the backbone in a similar fashion. In this fashion the Type II move geometry is determined. Any side chains attached to these rigid units are rigidly rotated when their parent unit is rotated.

The Type II rotation is chosen in the following manner.

25 Using the configurational bias prescription, the Hamiltonian is divided into uint and uext. uint is preferably 0, or alternatively is the torsional energy associated with the rigid unit of interest, while uext includes all remaining interaction energies. In the previous manner, uint determines of pint according to which are generated K' candidate \$\phi_0\$ rotation angles. Preferably K' is 1. Then the geometric constraints are solved for each candidate \$\phi_0\$. Typically, but not always, 6K', denoted K, possible backbone alterations are obtained. One of these is selected by \$p^{ext}\$, determined by:

$$p^{ext}(\phi_{0,k}) = \frac{\exp\left[-\beta u_0^{ext}(\phi_{i,k})\right]}{W^{ext}(\phi_{i,k})}$$

$$W^{ext}(\phi_{i,k}) = \sum_{k=1}^{K} \exp\left[-\beta u_0^{ext}(\phi_{i,k})\right]$$
(13)

5

uext includes all interactions not in uint, that is all other backbone and side chain interactions. Because these determinations occur in torsional angle space and change the volume element in that space, the Jacobian, determined by equation 35, of the selected Type II move is also needed as a weight in the acceptance probability for detailed balance. This acceptance probability for Type II moves is:

15

$$accept(curr-prop) = min[1, \frac{W^{nev}j^{nev}}{W^{old}J^{old}}]$$
 (14)

The weight and Jacobian of the reverse transformation
from the proposed to the current structure are also needed in
the acceptance probability for Monte Carlo detailed balance.
These quantities are determined as follows. Using the
proposed backbone structure just selected as the basis,
generate a set of K'-1 new φ₀ torsional angles according to

p^{int} and also include the current φ₀ in the set. Then solve
the geometric constraint to determine the permitted
alterations. The current configuration, since it exists,
must be among the permitted structures. From this set of
permitted structures determine W⁰¹⁴ per equation 13. Then
select the current configuration and compute the Jacobian J⁰¹⁴
per equation 35. This completes the determination or the
acceptance probability.

Proline is approximated. Proline is not subject to Type I moves. However, proline is subject to normal Type II moves, with its side chain bond to the amino nitrogen broken. The side chain thus moves rigidly with its backbone rigid unit as in normal Type II move. To compensate for the broken

bond approximation, the Co-N torsional energy amplitude in the proline backbone is set at approximately 5 kcal/mole. (By contrast the torsional energy in a typical amino acid of the Co-N bond is approximately 0.3 kcal/mole.) This invention is adaptable to other suitable approximations for proline. Alternatively, the proline side chain may be subject to alterations which preserve its cyclicity, such as for example, by an extension of the constraint scheme just described.

10

Program detailed description

The following describes the construction and use of a computer method and apparatus to perform the method of step 5. The listing of this code is included in a microfiche 15 appendix to this specification. Fig. 15 is a general view of the computer system and its internal data and program structures. To the left in Fig. 15 are the principal data structures of this method. Current structures 1701 contains the current structures of the N binders represented in memory 20 as described. Proposed structure 1702 contains working memory areas used to generate a proposed new structure for one binder peptide. Structures 1701 and 1702 would typically be stored in RAM memory of the computer system, RAM memory being memory directly accessible to processor fetches.

25 Stored structures 1703 contain similar memory representations of all the peptide structures generated, accepted, and selected for storage. This is typically stored on permanent disk file(s).

Candidate pharmacophore structures 1704 are input to the 30 programs from either a disk file of the display and input unit 1712. The identified candidate structures are used to determine the w'_{1.11} in Eqn 11.

Parameters 1705 comprises several parts. First, are all the AMBER atomic interaction definitions and parameters.

35 Second, are standard representations of the amino acids including component rigid units and atomic charge assignments. Third, are parameters controlling the run.

These further comprise, by example, values for K and K', the Type I/II move branching ratio, the number of moves made in the simulations run, the simulation total energy record, etc. The parameters would typically be loaded from disk file(s) into RAM memory for manipulation during a simulation run.

Unit 1712 includes display and input devices for monitoring and control. Depicted on the display are the total number of moves made in the current run and the course of the total energy, which is similar to that illustrated in 10 Fig. 9.

Processor 1711 is loaded with necessary programs prior to a simulation run and executes the programs to perform the simulation method. The general structure consists of main program 1706, structure modification program 1707, Type I and

- 15 II move generators 1708 and 1709, and subroutines 1710. The subroutines consist of common utility subprograms, such as for performing torsional rotations about bonds and computing interaction energies by the previous methods, and conventional library subprograms, such as for performing
- 20 input and output and finding random numbers. Any scientifically adequate random number generator can be used. A reference for random number generators is Press et al., Numerical recipes: the art of scientific computing, Cambridge, U.K., Cambridge University Press, (1986), chapter
- 25 7. The invention is equally adaptable to other program structures that will occur to those skilled in computer simulation arts.

The preferred embodiment of these structure is an Indigo 2 workstation from Silicon Graphics (Mountain View, CA).

- 30 Alternatively, any high performance workstation, such as products of Hewlett-Packard, IBM or Sun Microsystems, could be used. Preferably the data and program structures are coded in the C computer language. Alternatively any scientifically oriented language, such as Fortran, could be
- 35 used. Conventional subroutine and scientific subroutine libraries are used where appropriate.

The program components will be now 19. Fig. 16.

The program components 17. 18. and 19. Fig. 16.

The program components 17. 18. and 19. The peptide sequences of the The program components will be now describe.

The program to rigs. 1706.

With reference main program 1706. illustrates are input lengths and angles. Aramarere hond lengths Parameters interaction parameters 1802. N binders are input at step lend, and angles, acid definitions binders are bond lengths and parameters.

S Parameters interaction parameters. step 1803 creates initial for all step 1803 creates initial for all step 1803 creates initial records for all are input data.

are input at step input data. This input and linked to represent peptides.

structures are created and linked to represent peptides. charges interaction parameters amino acid definition step 1803 creates init records are input at step innut data. WO 96130849 structures from this input data. Rigid unit records for a Rigid unit records for a Rigid unit records for a Rigid unit represent pertides.

Structures from this input data. Rigid unit represent pertides are represent and linked to represent a rep The geometric structures of these peptides adding library of of the geometric structures of characteristic of the library of obtained from a prior characteristic of the library of obtained from a protocypical backbone characteristic of the library of the librar rigid units are created and linked to represent peptide either are nuit nv adding eide to represent peptides.

The geometric structures of these nuit nv adding eide n obtained from a prior run or are built by adding library of the library is to a prototypical backbone for the cx.c library is to a prototypical backbone for the cx.c library is to a binder. to a prototypical backbone appendix heading cx6C. CAR.

the binder. the microfiche appendix heading the found in the microfiche appendix heading the microfich found in the microfiche appendix heading the current structures are stored in the main stens of the beginning the main structures are structures are beginning the main structures are initial binder preparation for the beginning the main structures are as in preparation for the beginning the main structures. initial binder structures are stored in the current structure for the beginning the main steps of data areas in preparation for the beginning the main steps of the beginning the beginning the main steps of the beginning the begi nethod. 1804 begins the main loop of the simulation with for one of structure for one of step land begins the main loop of structure for one of step land begins the main loop of structure for one of step land begins the main loop of the simulation with l step 1804 begins the main loop of the simulation with for one of structure for one a proposed modified structure modification nrontan aproposed the generation of a proposed modification nrontan the generation of a proposed modification nrontan aproposed the proposed modification nrontan aproposed modification nrontan aproposed modification nrontan aproposed modification nrontan appropriate the generation of the simulation with the generation of a proposed modified structure program acceptance the binder peptides structure generation.

The binder peptides structure generation. the binder proposed structure generation, as previous as previous as previous as previous as previous as part of proposed (curr-zoron) is determined as previous as part accept (curr-zoron) as part accept (curr-zoron) accept (c As part of proposed structure generation as accepted at 1805 proposed structure will be accepted at accept court accept court accept court accept accept court accept accept accept accept accept accept court accept (curr-aprop) 1s determined as previously
will be accepted at 1805
will be accepted number
The proposed structure
nic prohability described. The proposed structure will be accepted at 18(

The proposed structure example, a random number for example, and the proposed structure based on this probability.

The proposed structure and the proposed structure and the proposed structure based on and 1 is denerated. based on this probability. For example, a random number than the acceptance and the proposed structure acceptance is less than the acceptance is less than the acceptance are perween 0 and 1 is generated, number is less than the acceptance are proposed if the random number is less than the acceptance are acceptance as acceptance. between 0 and 1 is generated, and the proposed structure is accepted then accepted if the proposed structure is accepted. Then accepted if the proposed structure is accepted. accepted if the random number is less than the accepted, then rhis

accepted if the proposed structure at step. 1806.

Probability.

Probability.

Probability. Probability. If the proposed structure is accepted, the nrong atomic nosition in the nrong atomic nosit is tested for sufficient one atomic position in the corresponding nosition in the test is met if at from the corresponding has been atomic position in the corresponding to the corresponding to the corresponding has been to the corresponding test is met differs from at least approximately no structure by a struct structure differs by at least approximately of at 1807 in the structure is distinct. current structure by at least approximately 0.2 Å. If the most of not proposed structure later analysis. Whether distinct or not proposed structure for later analysis. Whether distinct or not structure store for later structure at step 1808.

Structure accepted proposed structure at the accepted representation of the structure at the accepted current structure at the accepted proposed structure at the accepted prop 25 Ponding current structure at step 1808. at step 1809.
The simulation is tested by the onerator at starting on the controlled by the onerator at starting of the simulation of the controlled by the onerator at starting of the simulation of the controlled by the onerator at starting of the simulation of the controlled by the onerator at starting of the simulation of the controlled by the onerator at starting of the simulation of the controlled by the order of the order of the controlled by the order of the controlled by the order of the The simulation is tested by the operator at station 1712 completion can be controlled by proposed scruciure for later analysis. the accepted proposed structure at step land.

The accepted proposed structure at step land.

The accepted proposed structure at step land.

depending on display of run progress results. Alternatively, termination can be mechanically controlled. After completing a certain number of total moves after run energy equilibration, the moves being split between Types I and II according to the specified branching ratio, the run is terminated. The preferred number of total moves is 25,000, and the preferred Type I/II branching ratio is 4. Thus it is preferred to have 20,000 Type I and 5,000 Type II moves after equilibration per simulation run.

At step 1810, the stored structures are analyzed to 10 determine both the consensus pharmacophore structure and the structures of the remainder of the binders. In the preferred embodiment, atomic positions in the equilibrated stored structures for each peptide are averaged to obtain the 15 predicted geometric structure. The shared pharmacophore structure is obtained from the predicted structure of each peptide, again by averaging the shared position information for all peptides. Alternatively, before structure averaging, the structures generated for each binder can be clustered 20 into similar groups and the clusters for each peptide separately averaged. The clusters would represent alternative peptide folding patterns. It is anticipated that because preferred binders are short peptides constrained by disulfide bridges, any alternative foldings identified will 25 be structurally similar. The clustering can be done by the exemplary methods found in the previously referenced article Gordon et al. Fuzzy cluster analysis of molecular dynamics trajectories. Proteins: Structure, Function, and Genetics 14:249-264 (1992). For all analysis methods, the choice of 30 the preferred number of stored moves is adjusted to achieve adequate estimated statistical position errors. Further, preferably, the results of three runs are combined to achieve increased statistical confidence.

Other information is also output. Particularly

35 important is the course of the total energy for each peptide
and for all the peptides, and the intra-molecular, consensus,
and constraint components of the energies. These energy

components are used in determining whether a consensus pharmacophore has been found. As previously described, this is preferably done by insuring that $H_{consensus}$ is small compared to the total energy and is minimized by a particular 5 candidate pharmacophore. Also H_{MMR} must be relatively small.

Finally at 1811, all results are output in a form usable for the subsequent steps 6 and 7 of Fig. 1. For example, this may be a particular file format suitable for subsequent lead compound search by a database query.

Turning now to Fig. 17, structure modification program 1707 will be described. This is invoked from the main program at 1804. Upon entry, this program randomly picks one of the binder peptides at 1901 for which to generate a proposed structure and also picks which type of move to use 15 at 1902. This latter random choice is made according to an adjustable Type I/II branching ratio (preferably 4). For a Type I move, step 1903 picks a random amino acid side chain of the selected peptide, and step 1904 invokes the Type I move program. (Proline has no Type I moves.) For a Type II 20 move, step 1905 picks a random backbone bond between rigid units to rotate and also a random direction from the picked bond along which backbone rigid unit structure will be altered. Step 1906 invokes the Type II move program.

Figs. 18A and 18B illustrate the Type I move generator 25 1708, which is defined by equations 6 and 7. With reference first to Fig. 18A, the proposed structure of the selected peptide is created from its current structure by removing the selected side chain. All intra-molecular interactions are subsequently determined with respect to the proposed

- 30 structure absent side chain rigid units not yet regrown. K candidate new torsional angles for the next, i'th, rigid unit to add are generated by p_i at 2002. Preferably K is between 10 and 100. Generation of these angles uses the conventional rejection method referenced in Press et al. at § 7.3. The
- 35 weight w_i^{ext} and p_i^{ext} are determined for each of these candidate angles. This requires the rigid unit to be added to be rotated to the candidate angle using the previous

rotation method. Candidate interaction energy is determined from candidate interatomic distances resulting from the candidate rotation. One of the candidate angles is probabilisticly selected at 2003 and the rigid unit added 5 back at this torsional angle at 2004. If there are more units to add, which is tested at 2005, these steps are repeated. If not, the acceptance weight W^{new} is determined as the product of the w_i^{ext} at 2006. Lastly the old weight is determined at 2007. From the weights the move acceptance 10 probability is found for use at 1805.

Fig. 18B details the determination 2007 of Wold, the weight for the reverse move from the proposed to the current side chain structure. Temporarily the proposed structure is used as a basis for energy determination at 2008, and then 15 the current structure is restored at 2016, when this process is finished. The proposed side chain is removed at 2009 for regrowth rigid unit by rigid unit as in Fig. 18A. For the next, i'th, rigid unit to be added back, K-1 candidate angles. are generated according to pint at 2010 with the current value 20 of that angle for the K-th candidate at 2011. As previously, the weight w, ext is determined for these candidate angles at 2012. The rigid unit is added back at the current, K-th, angle at 2013. If there are more units to add, tested at 2014, these steps are repeated. If not, the acceptance 25 weight Wold is determined as the product of the w.ext at 2006. Figs. 19A and 19B illustrate Type II move generator

Figs. 19A and 19B illustrate Type II move generator 1709, which is defined by equation 13 and 14 and the concerted rotation geometric constraints. With reference to Fig. 19A, K' candidate new torsional angles for the selected 30 backbone bond are generated by p^{int} using the rejection method. Preferably K' is 1. Torsional rotations about adjacent backbone bonds, in the selected direction along the backbone, permitted by the concerted rotation constraints are determined from the roots of equation 34 at 2102. Equation 34 depends on intermediate variables obtained from equations 25, 29, and 32 and determined in that order. The roots are simply found by searching the interval [-π,π] in 0.04°

increments. When a root is located in a 0.04° segment, it is refined with the bisection method referenced in Press et al. at § 9.1. It is expected on the average that six K' solutions will be found. If no roots are found at 2103, the 5 candidate rotation is impossible and this move is skipped. If solutions exist, next, at 2104, pext and Wnex are determined. Using the described rotation method, the backbone rigid units are rotated (with consequent spatial displacement of 4 units) to a candidate torsional angle solution about their mutual

- 10 bonds. Additionally, any side chains attached to backbone rigid units are rigidly rotated using the same method. Having made these rotations, candidate interatomic distances and candidate interaction energies can be determined and used to obtain p^{ext} for this candidate solution. One of the
- 15 candidates is probabilisticly selected at 2104, and the backbone and any side chains are rotated according to this candidate into the proposed structure. The Jacobian of this transformation is determined at 2106 by equation 35. Lastly the old acceptance weight and Jacobian are determined at
- 20 2107. From the weights and Jacobians the move acceptance probability is found for use at 1805.

Fig. 19B details the determination 2107 of W^{old} and J^{old} for the reverse move from the proposed to the current side chain structure. Temporarily the proposed structure is used 25 as the basis for energy determination at 2008, and the current structure is restored at 2016, when this process is finished. At 2109, a set of K'-l candidate torsional angles is generated for the selected backbone bond according to p^{int} using the rejection method and the current torsional angle is 30 added to this set. If as preferred, K' is 1, this step results in a set with only the current angle. At 2111, similarly to 2102, the permitted torsional rotations about adjacent backbone bonds are determined from the equations expressing the concerted rotation constraints. Special care 35 is taken to ensure that the original conformation is found by

35 is taken to ensure that the original conformation is found by the root finding procedure. In particular, the search interval is centered on the known original ϕ_1 and is made as

small as necessary to isolate the root, which may be as small as 0.004° or smaller. The current structure must be among these solutions, since it exists. Select it at 2112. W^{old} is computed from the candidate angle solution, making the 5 candidate rotations and determining candidate interactions. Also the Jacobian, J^{old}, of the transformation is computed from the proposed to the current structure.

5.8. CONSENSUS STRUCTURE TEST

- Having selected a candidate pharmacophore and determined a best possible consensus structure and best possible structures for the remainder of the binder molecules, the consensus test, step 6, tests whether a consensus structure has actually been found. A consensus pharmacophore structure
- 15 consists of a spatial arrangement of chemically similar groups shared by all the N binders to high accuracy. Since an actual pharmacophore exists, the N specifically binding members of the screened libraries will share the actual structure. However, the remainder of binder molecules will
- 20 share no other similar structures to such a high accuracy.

 Therefore, a structure consensus of the N binders is possible only if the candidate pharmacophore is the actual physical pharmacophore responsible for the actual binding. If the candidate selected relates to other parts of the binder
- 25 molecules, no structure consensus will be found. Further, if the Monte Carlo determination attempts to impose a consensus on parts of the binder molecules that do not share structure, an inconsistent overall structure will be obtained for the remainder of the binder molecules.
- Therefore, two preferred consensus tests are applied:
 one test asks whether a consistent candidate pharmacophore
 has been obtained, and a second test asks whether consistent
 structures have been obtained for the remainder of the binder
 molecules. Both tests have a preferred absolute and a less
 preferred relative version.

There are two portions for the first test. First, are all the consensus pharmacophore distances obtained in the N

```
binders within at least a specified distance, preferably
                                                                                                                                                                                                          binders within at least a specified distance, preference on announce of the stance of 
                                                                                                                                                                                                    approximately of each other; second, is the energy (e.g., less than at most annroximare)
                                                                                                                                                                                              consensus energy, Konsensus, relatively small compared to the total molecular energy) as determined by the
                                                                                                                                                                              5 5-10; molecular energy (e.g., less than at most approximate the folial molecular energy) as determined by the
                                                                                                                                                                       There are also two portions of the second test.

Can the intramolecular portions of the second test.

Can the intramolecular portions of the second test.

Confirmed by additional distance by the Monte Carlo
                                                                                                                                                                                                           There are also two portions of the second test.
                                                                                                                                                                   The intramolecular distances predicted by the Monte Carlo method urilized by the Monte Carlo method urilized address.
                                                                                                                                            To Second, be confirmed by additional distance measurements or more of these distance of the more of t
                                                                                                                                                       Second, since the Monte Carlo method utilizes distance can be ignored, one of these head; or the near; or the
                                                                                                                                                  constraints
measurement previously measured, one or more of these checked against that measured and the predicted roleran.
                                                                                                                                            distance checked against that measured distance or are nicrance measured distance.
                                                                                                                                      for these tests against that measured distance.
e. v. addistance agreements of at least specified

each binder.
                                                                                                                15 distances, e.g., approximately 0.5 Å, in each binder.

The two preferred tests have been described in the
                                                                                                                                                       The two preferred tests have been described in the
                                                                                                                    absolute version as requiring checks against absolute

The value of the version of the absolute

The value of the absolute
                                                                                                               tolerances. Alternatively, the values of the pharmacophore
                                                                                                          distances.

Alternatively, the values of the pharmacopn distances distances and the distances can t
                                                                                   distance differences among the binders, Remains and the Dossible candidate Dharmacophores.
                                                                                               accumulated for all the predicted and measured distances can be minimizing these
                                                                                         the candidate for all the possible candidate pharmacophor the selected being that one minimizing these will have
                                                                                   the candidate selected being that one minimizing these of the pharmaconhore the
                                                                              departures.

Minimum Values for the differences of the pharmacophore

the minimum Value for He will have
                                                      25 distances in the differences of the minimum values of the minimum values of the minimum values of the differences of predicated from
                                                                  the minimum values of the differences of predicated from
                                                                                     This invention is adaptable to other tests that evaluate obtained for the
                                                   the Consistency of the consensus structure obtained for the And the Accuracy of the consensus of the contract of the cont
                           the consistency of the obtained for the remainder of the binder molecules.
Having 5.9.

Jetion of step at step 1 with a target of interest, upon
         pharmacophore step 6 or rig. 1 a might resolution support into structure has been determined as well as this hinder mentides. This h
     supporting structure nas been determined as well binder peptides.
```

```
resolution structure is as a drug that will hind to the compounds for use as a
                                                                                                                       resolution structure is a drug that will bind to the original compounds for use as a
                                                                                                                                                                                                                                                            interest. lead compounds are determined, that of interest. one or more specification with the determined that the specification with the determined that the specific at ion with the spec
                                                                                                                                          Thus one or more lead compounds are determined the determined the or more specification with the determination of the a pharmacophore structure.

S share a pharmacophore structure of the consensus of the consen
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         This determination can be
                                                                                                                                                                                                                   consensus pharmacophore of several methods: or of chemical drug compounds or of chemical preferably done of porential drug compounds or of chemical drug com
WO 96/30849
                                                                                                                                                                                                                                         preterably cone by one of several methods: by a search chemical compounds or of chemical drug compounds or of chemical crandard name rile from the crandard name of potential crandard name crandard n
                                                                                                                                                                                                                   structures (e.g., London, Frankfurt, Germany or Chicago) and structures Ltd., Frankfurt, Germany or Chicago (Bielstein Information, Frankfurt, Germany or Chicago)
                                                                                                                                                                                                      Snare a pharmacophore structure.
                                                                                                                                                                                                                                                                                                  Publications Ltd., London, Frankfurt, Germany of identify (CAS, Columbus, OH) to identify (CAS, Columbus, Columbus, Columbus, OH) to identify (Eielstein Information, CAS, Columbus, OH)
                                                                                                                                                                                                                                                                                                                     (Bielstein Information, (CAS, Columbus, Specification, the Charmacophore Specification, the Charmacophore that compounds that contain
                                                                                                                                                                                                                                                                                                                                    the Chemical Registry the pharmacophore include the compounds that compound to a known lead compound to modification of a known lead compound to include the compounds that compounds the compo
                                                                                                                                                                                                                                                                                                                                                       compounds that contain the pharmacophore specification of a known lead compound to include and not modification of a conformation of a con
                                                                                                                                                                                                                                                                                                                     modification of a known lead compound to include the synthesizing a de novo by synthesizing a de novo by synthesizing a de novo by synthesizing a known lead by synthesizing a perification.

modification of a known lead compound to include the novo include the n
                                                                                                                                                                                                                                                                                                                                                                                                pharmacophore specification; by synthesizing a de novo or by synthesizing 
                                                                                                                                                                                                                                                                                                                                                                                                                                modification of binders nore attractive for use as a druo (e.g., no isolated in binder nore attractive for use as render the binder nore
                                                                                                                                                                                                                                                                                                                                                                                                                structure containing the pharmacophore specification; (e.g., the pharmacophore molecule (e.g., the pharmacophore structure containing to the the pharmacophore structure containing to the the pharmacophore structure containing the pharmacophore modern and pharmacophore structure containing the pharmacophore specification; (e.g., the pharmacophore specification) and the pharmacophore specification of pinders to the pharmacophore specification of pinders to the pharmacophore specification of pinders to the pharmacophore specification of pinders in the pharmacophore specification of pi
                                                                                                                                                                                                                                                                                                                                                                                                                                                   isolated in step 2) outside of the pharmacophore structure to the pharmacophore at tractive for use as a drug desired in more attractive ability to achieve desired in render half-life. solubility
                                                                                                                                                                                                                                                                                                                                                                                                                render the binder more attractive for use as a drug desired in solubility to achieve desired in a increase half-life. solubility ability to achieve desired in the binder more attractive for use as a drug desired in achieve desired in a chieve des
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          localization). Queries are based not only on chemical on precise decomposition. A parabase search but also on precise decomposition but also decomposition but al
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  Database search queries are pased not only on control on ast precise geometric parabase search but also on precise rely on ast information but also approaches rely on ast control of the committee of approaches approaches
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            information. find matching J. Medicinal Chemistry, vol. 35. To searching in drya design, searchi
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       rmarion puc arso on precise geometric database computer based approaches rely on database
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       searching in druc design, herein incorporated by reference.

searching in druc design, herein incorporated by 2145-54 (1992).
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             Searching in drya design, herein incorporated by reference.

Searching in drya design, herein 2-D and 3-D darahages of 2145-54 (1992), for searching 2-D and 3-D arahages for searching 2-D and 3-D darahages of Existing methods
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     vivo localization).
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                2145-54 (1992), which is nerein incorporated by reference american incorporated by reference are applicable to this step.

Existing methods for searching to this step.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Lederle of American
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             compounds are applicable to this step. Lederle of America nas pioneered molecular has pioneered molecular of darana and trend-vectors of darana and cyanamid (pearl River, 3D searching and trend-vectors) has pioneered molecular are darana pioneered molecular has been pioneered molecular has pio
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            Cyanamid (Pearl River, New York) has Pioneered molecular of databases.

Cyanamid (Pearl River, New York) and trend-vectors of enhanced and other research droups have enhanced and other research croups have enhanced and other research croups have enhanced and other research croups have enhanced and other research droups have enhanced and other research croups have enhanced and other research droups have enhanced and droups 
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               Snape-searching, 3D searching and trend-vectors of databases

Snape-searching, 3D searching research groups have enhanced trend-vectors and trend-vectors have enhanced tr
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               Commercial vendors and other research groups have enhanced [San Molecular Design Ind. (San Molecular Design Of MACSS-3D, Molecular University of et al., University of searching capabilities Lauri, G. et al., University Lauri, 
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  Bearching capabilities Lauri CHEM-X. Chemical Design To Leandro (Berkelev. CA):
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               Leandro, CA); CAVEAT, CA); CREM-X, Chemical Design, Inc.

Leandro, CA); (Berkeley, CA); CREM-X, Chemical Design, Inc.

(Mahwah, N.J.)).
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           (Wahwah, W.J.)).
```

The pharmacophore structure determined in this invention is adaptable to any of these methods and sources of chemical database searching and to the enumerated non-database methods. Output will be lead compounds suitable for drug 5 design. An important aspect of this invention is that the high resolution pharmacophore structure will lead to highly targeted leads. Lower resolution structures result in a geometric increase in the number of lead compound query matches. Example 1 illustrates this effect.

10

5.10. APPENDIX: CONCERTED ROTATION

Since the preferred molecules under consideration are conformationally constrained by disulfide bridge(s), a Monte Carlo move that preserves this constraint is required. The 15 "concerted rotation" scheme used for alkanes can be extended to allow rotation of the torsional angles in conformationally constrained peptides. This appendix describes this extension. Dodd et al. (1993) discusses the original, restricted method. (The essential extensions are expressed in equations 27, 28, and 34.) This method is directly applicable to the cyclic residue of proline, and an alternative embodiment of this invention would thermally perturb proline with a move of similar geometric constraints.

Fig. 14 illustrates the geometry under consideration.

- 25 Illustrated backbone 1600 is a poly-glycine 7-mer. Rigid unit positions are indicated generally by black circles as at 1601 with incoming bonds generally as at 1602. The torsional rotations ϕ_0 to ϕ_6 are about bonds 1602 to 1608, respectively, between sequential, adjacent rigid units. The rigid unit
- 30 position vectors \underline{r}_0 to \underline{r}_6 , illustrated as vectors 1610 to 1616, respectively, define the position of these sequential rigid units with respect to a laboratory coordinate system with origin 1609. A C_a rigid unit (B unit) is illustrated in box 1630, and an amide bond (C unit) in box 1631.
- To formulate this method, let us consider rotating about seven torsional angles, which will displace the root positions and rotate four rigid units, rotate up to three

additional ones, and leave the rest of the peptide fixed. The root position of a rigid unit is the C_o position for a B unit, the C position for a C unit, the C position for a CH₂ unit, and the S position for the S unit in cystine. If unit 5 is a C unit, however, \underline{r}_{i} is defined to be the backbone amino nitrogen position of that unit. For each unit, let us define θ_{i} to be the fixed angle between the incoming and outgoing bonds. Thus, $\theta_{i} = 0$ for a C unit, and $\theta_{i} = 70.5^{\circ}$ for all others.

The method leaves the positions \underline{r}_i of units $i \leq 0$ or $i \geq 5$ fixed. The torsion ϕ_o is changed by an amount $\delta \phi_o$. The values of ϕ_i , $1 \leq i \leq 6$ are then determined so that only the positions r_i of units $1 \leq i \leq 4$ are changed.

The method requires several definitions to present the 15 solution for the new torsional angles. The bond vectors are defined to be the difference in position between unit i and unit i - 1, as seen in the coordinate system of unit i:

$$I_{z} = \mathcal{I}_{z}^{(i)} - \mathcal{I}_{z-1}^{(i)}. \tag{15}$$

20

Bond vectors \underline{l}_1 to \underline{l}_5 are illustrated in Fig. 14 at 1620 to 1624, respectively. The length and orientations of the \underline{l}_i are determined by rigid unit structure and the length and angle AMBER parameters for bonds between atom types. The coordinate system of i is such that the incoming bond is along the \hat{x} direction. Thus $\underline{l}_i = l_i \hat{x}$ if atoms r_i and r_{i-1}

are directly bonded to each other and has x- and y-components otherwise. Here \hat{x} is a fixed unit vector along the x

direction. Now define a rotation matrix that transforms from the coordinate system of unit i+1 to unit i

$$T_{i} = \begin{pmatrix} \cos\theta_{i} & \sin\theta_{i} & 0\\ \sin\theta_{i}\cos\phi_{i} & \cos\theta_{i}\cos\phi_{i} & \sin\phi_{i}\\ \sin\theta_{i}\sin\phi_{i} & -\cos\theta_{i}\sin\phi_{i} & -\cos\phi_{i} \end{pmatrix}$$
(16)

5

The positions of the units in the frame of unit 1 are, thus, given by:

 $\begin{array}{rcl}
\mathbf{I}_{1}^{(1)} &= \mathbf{1}_{1} \\
\mathbf{I}_{2}^{(1)} &= \mathbf{1}_{1} + \mathbf{T}_{1} \mathbf{1}_{2} \\
\mathbf{I}_{3}^{(2)} &= \mathbf{1}_{1} + \mathbf{T}_{1} \left(\mathbf{1}_{2} + \mathbf{T}_{2} \mathbf{1}_{3} \right) \\
\mathbf{I}_{4}^{(1)} &= \mathbf{1}_{1} + \mathbf{T}_{1} \left(\mathbf{1}_{2} + \mathbf{T}_{2} \left(\mathbf{1}_{3} + \mathbf{T}_{3} \mathbf{1}_{2} \right) \right)
\end{array} \tag{17}$

15

Further define the matrix that converts from the frame of reference of unit 1 to the laboratory reference frame

$$T_1^{lab} = [\cos\psi I + nn^T (1 - \cos\psi) + M\sin\psi] A.$$
 (18)

20

where

$$M = \begin{pmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{pmatrix}$$
 (19)

25

and

$$D = \frac{\hat{x} \times I}{|\hat{x} \times I|}$$

$$\cos \psi = \frac{I \cdot \hat{x}}{|I||\hat{x}|}$$

$$\sin \psi = \frac{|(I \times \hat{x})|}{|I||\hat{x}|},$$

where \underline{r} is the axis of the bond coming into unit 1. The matrix A is a rotation about \hat{x} and is defined so that

 $_{5}$ Al, = Δ r :

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c & -s \\ 0 & s & c \end{pmatrix} \tag{20}$$

10

where

$$c = (l_{1y}\Delta r_y + l_{1z}\Delta r_z) / (\Delta r_y^2 + \Delta r_z^2)
s = (-l_{1z}\Delta r_y + l_{1y}\Delta r_z) / (\Delta r_y^2 + \Delta r_z^2).$$
(21)

15

Here $\Delta R = A[T_1^{lab}]^{-1}(x_1-x_0)$ if unit 0 is a C unit. Otherwise,

 $\Delta \underline{r} = \underline{l}_1.$

The method proceeds by solving for ϕ_1 , $2 \le i \le 6$, analytically in terms of ϕ_1 . Then a nonlinear equation is solved numerically to determine which values of ϕ_1 , if any, are possible for the chosen value of ϕ_0 .

The derivation proceeds in the coordinate system of unit 25 1, after it has been rotated by the chosen ϕ_0 . Define

$$\underline{\mathbf{r}} = \underline{\mathbf{r}}_{5}^{(1)} - \underline{\mathbf{l}}_{1} = [\underline{\mathbf{r}}_{1}^{1ab}]^{-1} (\underline{\mathbf{r}}_{5} - \underline{\mathbf{r}}_{0}) - \underline{\mathbf{l}}_{1}. \tag{22}$$

If $\theta_3 \neq 0$ and $\theta_5 \neq 0$, one can see from Fig. 14 that the distance between unit 3 and unit 5 is known and equal to

$$q_1^2 = \frac{(l_{4x}\cos\theta_4 - l_{4y}\sin\theta_4 + l_{5x})^2 + (l_{4x}\sin\theta_4 + l_{4y}\cos\theta_4 + l_{5y})^2}{(l_{4x}\sin\theta_4 + l_{4y}\cos\theta_4 + l_{5y})^2}$$
 (23)

But this distance can also be written as 35

WO 96/30849

$$q_1^2 = |x - T_2|_3|^2$$

 $x = T_1^{-1} \pm - 1_2$. (24)

5 Equating these two results, two values of ϕ_2 are possible

$$\phi_2^I = \arcsin(c_1) - \arctan(x_y/x_z) - H(x_z)$$

$$\phi_2^{II} = \pi - \arcsin(c_1) - \arctan(x_y/x_z) - H(x_z),$$
(25)

10

with

$$H(x) = \begin{cases} 0, & x > 0 \\ \pi, & x < 0 \end{cases}$$
 (26)

15 The constant c_1 is given by

$$C_{1} = \begin{cases} \frac{q_{1}^{2} - x^{2} - l_{3}^{2} + 2x_{x}(\cos\theta_{2}l_{3x} + \sin\theta_{2}l_{3y})}{-2(\sin\theta_{2}l_{3x} - \cos\theta_{2}l_{3y})(x_{y}^{2} + x_{z}^{2})^{1/2}}, & \theta_{3} \neq 0, \theta_{5} \neq 0 \end{cases}$$

$$C_{1} = \begin{cases} \frac{l_{3x} + l_{4x} + l_{5x}\cos\theta_{4} - x_{x}\cos\theta_{2}}{\sin\theta_{2}(x_{y}^{2} + x_{z}^{2})^{1/2}}, & \theta_{3} = 0, \theta_{5} \neq 0 \end{cases}$$

$$C_{1} = \begin{cases} \frac{(\mathcal{I}_{5} - \mathcal{I}_{2}) \cdot (\mathcal{I}_{6} - \mathcal{I}_{5}) / l_{6} - l_{5} - l_{4x}\cos\theta_{4} - x_{x}(\cos\theta_{2}l_{3x} + \sin\theta_{2}l_{3y})}{(\sin\theta_{2}l_{3x} - \cos\theta_{2}l_{3y})(x_{y}^{2} + x_{z}^{2})^{1/2}}, & \theta_{3} = 0, \theta_{5} = 0 \end{cases}$$

$$\frac{l_{3x}\cos\theta_{4} - x_{x}(\cos\theta_{2}l_{3y} + \sin\theta_{2}l_{3y})}{(\sin\theta_{2}l_{3x} - \cos\theta_{2}l_{3y})(x_{y}^{2} + x_{z}^{2})^{1/2}}, & \theta_{3} = 0, \theta_{5} = 0 \end{cases}$$

$$(27)$$

30

where \underline{x} is given by Eqn. 24 if $\theta_s \neq 0$, and $\underline{x} = \underline{T_1}^{-1} [\underline{T_1}^{lab}]^{-1} (\underline{r_6} - \underline{r_5})/l_6$ if $\theta_s = 0$. Clearly for there to be a solution $|c_1| \leq 1$. The last three equations for c_1 were determined by conditions similar to equating Eqns. 23 and 24. For $\theta_3 = 0$, $\theta_5 \neq 0$, the

x component of $\underline{r}_5^{(3)} - \underline{r}_3^{(3)}$ is known to be equal to $(l_{4x} + l_5\cos\theta_4)$. For $\theta_3 \neq 0$, $\theta_5 = 0$, the x component of $\underline{r}_5^{(5)} - \underline{r}_3^{(5)}$ is known to be equal to $l_{5x} + l_{4x}\cos\theta_4$. For $\theta_3 = 0$, $\theta_5 = 0$, the angle between $\underline{r}_3 - \underline{r}_2$ and $\underline{r}_6 - \underline{r}_5$ is known to be equal to θ_4 . To determine ϕ_3 two expressions for $|\underline{r}_5 - \underline{r}_4|^2$ are again equated to determine that:

$$c_{2} = \frac{l_{5}^{2} - y^{2} - l_{c}^{2} + 2y_{x}(\cos\theta_{3}l_{4x} + \sin\theta_{3}l_{4y})}{2(\sin\theta_{3}l_{4x} - \cos\theta_{3}l_{4y})(y_{y}^{2} + y_{x}^{2})^{1/2}}$$
(28)

10

15

where $y = T_2^{-1}(T_1^{-1} t - l_2) - l_3$. Again, $|c_2| \le 1$ for there to be

a solution.

If $\theta_s \neq 0$, the value of ϕ_s can now be determined from:

20

$$\mathcal{I}_{5}^{(1)} = \mathcal{I}_{4}^{(1)} + T_{1}T_{2}T_{3}T_{4} \mathcal{I}_{5}.$$
 (30)

Defining

$$Q_3 = T_3^{-1}T_2^{-1}T_1^{-1} \{T_1^{1ab}\}^{-1} (\underline{r}_5 - \underline{r}_4).$$
 (31)

25

the equations that define ϕ_4 are given by

$$q_{3y} = \cos\phi_4 (\sin\theta_4 l_{5x} - \cos\theta_4 l_{5y})$$

 $q_{3z} = \sin\phi_4 (\sin\theta_4 l_{5x} - \cos\theta_4 l_{5y})$ (32)

30

This is a successful rotation if the position of \underline{r}_{ϵ} is successfully predicted. That is, the equation

$$\underline{\Gamma}_{6}^{(1)} - \underline{\Gamma}_{5}^{(1)} = \underline{T}_{1} \underline{T}_{2} \underline{T}_{3} \underline{T}_{4} \underline{T}_{5} \underline{I}_{6} = [\underline{T}_{1}^{1ab}]^{-1} (\underline{\Gamma}_{6} - \underline{\Gamma}_{5}).$$
 (33)

must be satisfied. Consider the x-component, which implies

$$F_{5}(\phi_{1}) = \begin{cases} (\mathcal{L}_{6}^{(1)} - \mathcal{L}_{5}^{(1)})^{T} T_{1} T_{2} T_{3} T_{4} \hat{\mathcal{L}} - (l_{6x} \cos \theta_{5} + l_{6y} \sin \theta_{5}) = 0, & \theta_{5} \neq 0 \\ (\mathcal{L}_{4} - \mathcal{L}_{3}) \cdot (\mathcal{L}_{6} - \mathcal{L}_{5}) - l_{4} l_{6} \cos \theta_{4} = 0, & \theta_{3} \neq 0, \theta_{5} = 0 \end{cases}$$

$$|\mathcal{L}_{6} - \mathcal{L}_{4}| - [(l_{6x} + l_{5x})^{2} + l_{5y}^{2}]^{1/2} = 0, & \theta_{3} = 0, \theta_{5} = 0$$

$$(34)$$

must be satisfied if the rotation is successful. The equations for the case θ_s = 0 clearly express the geometric conditions required for a successful rotation.

5

Eqn. 34 is the nonlinear equation for ϕ_1 because ϕ_2 , ϕ_3 , and ϕ_4 are determined by Eqns. (25), (29), and (32) in terms of ϕ_1 . This equation has between zero and four values for each value of ϕ_1 , however, due to the multiple root character of Eqns. (25) and (29). The equation is solved by searching the region $-\pi < \phi < \pi$ for zero crossings. The search is in increments of $\sim 0.04^{\circ}$. These roots are then refined by a bisection method.

The transformation from φ₁, 0 ≤ i ≤ 6 to the new solution which is constrained to change only r₁, 1 ≤ i ≤ 4 actually implies a change in volume element in torsional angle space. This change in volume element is the reason for the appearance of the Jacobian in the acceptance probability.
The Jacobian of this transformation is calculated in Dodd et al. (1993) at pp. 991-93. It is slightly different here since root position r₅ is not necessarily the head position. The Jacobian is given by.

$$J = \frac{1}{|\text{detB}|} \tag{35}$$

where the 5 x 5 matrix B is given by $B_{ij} = [\underline{u}_j \times (\underline{r}_5 - \underline{h}_j)]_i$ for i

35 \leq 3 and $B_{ij} = [\underline{u}_j \times (\underline{r}_6 - \underline{r}_5)/|\underline{r}_6 - \underline{r}_5|]_{i-3}$ for i = 4,5. Here \underline{h}_i = \underline{r}_i , except that \underline{h}_5 is the head position even if $\theta_5 = 0$, and \underline{u}_i is the incoming bond vector for unit i.

Repeated application of the concerted rotation may lead to a slightly imperfect structure, due to numerical precision errors. In an alternative embodiment, peptide geometry would be restored to an ideal state by application of the Random 5 Tweek algorithm after several thousand moves (Shenkin et al., 1987, Biopolymers 26:2053-85).

The invention is further described in the following examples which are in no way intended to limit the scope of the invention.

10

6. EXAMPLES

6.1. RELATION BETWEEN EFFECTIVENESS OF POTENTIAL DRUG IDENTIFICATIONS AND PHARMACOPHORE GEOMETRIC TOLERANCE

Searches of a drug library well known to medicinal chemists, the Standard Drugs File (Derwent Publications Ltd., London, England), illustrate the geometric increase in the number of compounds found (and thus decrease in expected effectiveness of identification of potential drugs) as pharmacophore geometric tolerance is increased. Table 4 tabulates the results.

20

Table 4

	5HT3 (5 Hydroxytryptophan)			
	· Tolerance (Å)	Number of drug compounds		
25	2.0	64		
	1.0	35		
	0.5	27		
	0.25	12		
30	0.10	1		

	Dopamine		
	Tolerance (Å)	Number of drug compounds	
	2.0	188	
5.	1.0	185	
	0.5	60 ⁻	
	0.25	48	
	0.10	5	

10 The pharmacophores are two well known neurotransmitters, 5-hydroxytryptophan and dopamine. As the tolerance of one distance in the pharmacophore structure is decreased from 2.0 to 0.1 Å, the number of compounds retrieved from the database is listed. The advantage of achieving pharmacophore

15 resolution better than approximately 0.25 Å is clear.

If the tolerance of three distances were involved, the expected number of compound retrieved would be the cube of these numbers. For the dopaminergic pharmacophore, the number of lead compounds would decrease from over 6.5x10° to 20 about 125 as three tolerances were decreased from 2.0 Å to 0.1 Å.

This example illustrates the geometric increase in the number of leads identified as pharmacophore geometry is less well defined. It thus a very preferred aspect of this invention that the computational method results in determining pharmacophore structure accurate to at least approximately 0.25 to 0.30 Å. Thus an exponentially large improvement in lead compound selection for drug design can be expected to result from this invention.

30

6.2. EXPRESSION AND PURIFICATION OF TARGET PROTEINS

Target molecules that are proteins, for example ras, raf, vEGF and KDR, are expressed in the *Pichia pastoris*sexpression system (Invitrogen, San Diego, CA) and as glutathione-S-transferase (GST)-fusion proteins in *E. coli* (Guan and Dixon, 1991, Anal. Biochem. 192:262-267).

affinity matrix and recovered by washing, while the GST tag remains bound to the matrix. Milligram quantities of recombinant protein per liter of *E. coli* culture are expected to be obtainable in this manner.

5

6.3. SYNTHESIS AND SCREENING OF POLYSOME-BASED LIBRARIES ENCODING RANDOM CONSTRAINED PEPTIDES OF VARIOUS LENGTHS

6.3.1. PREPARATION OF DNA TEMPLATES

DNA libraries with a high degree of complexity are made as two components: an expression unit, and a semi-random (or degenerate) unit. The expression unit has been synthesized chemically as an oligonucleotide (termed T7RBSATG), and contains the promoter region for bacteriophage T7 RNA

- polymerase, a ribosome binding site, and the initiating ATG codon. The random region, also synthesized as an oligonucleotide (termed MMN6) contains a region complementary to the expression unit, the antisense version of the codons specifying Cys-X₆-Cys, and a restriction site (BstXI). The
- 20 library is constructed by annealing 100 pmol of
 oligonucleotide T7RBSATG [having the sequence
 5'ACTTCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCCAGAAAT
 AATTTTGTTTAACTTTAACTTTAAGAAGGAGATATACATATGCAT3'
- (SEQ ID NO:2)); and oligonucleotide MNN6 [having the sequence 5'CCCAGACCCGCCCCCAGCATTGTGGGTTCCAACGCCCTCTAGACA[MNN], ACAATG TATATCTCCTTCTT3' (SEQ ID NO:3); M = A or C , N = G, A, T, or C], and extending the DNA in a reaction mixture containing 10-100 units of Sequenase (United States Biochemical Corp., Cleveland, OH), all four dNTPS (at 1 mM), and 10 mM
- dithiothreitol for 30 min at 37°C. The extended material is then digested with BstXI, ethanol precipitated and resuspended in water. This fragment of DNA is then ligated via the BstXI end to a 250 base pair (bp), PCR-amplified Glycine-Serine coding fragment derived from gene III of M13
- bacteriophage DNA. The gene III fragment has been amplified by use of two primers, respectively termed FGSPCR [having the sequence 5'TCGTCTGACCTGCCTCAACCTCCCCACATGCTGGCGGCGGCTCTGGT3'

The cDNAs of these target proteins are cloned in the Pichia expression vectors pHIL-S1 and pPIC9 (Invitrogen). Polymerase chain reaction (PCR) is used to introduce six Histidines at the carboxy-terminus of these proteins, so that 5 this His-tag can be used to affinity-purify these proteins. The recombinant plasmids are used to transform Pichia cells by the spheroplasting method or by electroporation. Expression of these proteins is inducible in Pichia in the presence of methanol. The cDNAs cloned in the pHIL-S1 10 plasmid are expressed as a fusion with the PHO1 signal peptide and hence are secreted extracellularly. Similarly cDNAs cloned in the pPIC9 plasmid are expressed as a fusion with the α -factor signal peptide and hence are secreted extracellularly. Thus, the purification of these proteins is 15 simpler as it merely involves affinity purification from the growth media. Purification is further facilitated by the fact that Pichia secretes very low levels of homologous proteins and hence the heterologous protein comprises the vast majority of the protein in the medium. The expressed 20 proteins are affinity purified onto an affinity matrix containing nickel. The bound proteins are then eluted with either EDTA or imidazole and are further concentrated by the use of centrifugal concentrators.

As an alternative to the Pichia expression system, the 25 target proteins are expressed as glutathione-S-transferase (GST) fusion proteins in E. coli. The target protein cDNAs are cloned into the pGEX-KG vector (Guan and Dixon, 1991, Anal. Biochem. 192:262-267) in which the protein of interest is expressed as a C-terminus fusion with the GST protein.

30 The pGEX-KG plasmid has an engineered thrombin cleavage site at the fusion junction that is used to cleave the target protein from the GST tag. Expression is inducible in the presence of IPTG, since the GST gene is under the influence of the tac promoter. Induced cells are broken up by sonication and the GST-fusion protein is affinity purified onto a glutathione-linked affinity matrix. The bound

protein is then cleaved by the addition of thrombin to the

(SEQ ID NO: 4)], and RGSPCR [having the sequence 5'ATCAAGTTTGCCTTTACCAGCATTGTGGAGCGCGTTTTCATC3'
(SEQ ID NO:5)], and Taq DNA polymerase (Gibco-BRL). The amplified DNA (250 bp) was cut with BstXI to yield a 200 bp fragment that has been gel purified. The 200 bp fragment is then ligated to the random peptide coding DNA fragment. This DNA specifies the synthesis of a peptide of the sequence Met-His-Cys-(X)₆-Cys- (SEQ ID NO:6) fused to the Gly-Ser rich region of the Ml3 gene III protein. The Gly-Ser rich domain is thought to behave as a flexible linker and assist in presentation of the random peptide to the target molecules.

To make constrained random peptides of different lengths, oligonucleotides are made that are similar to MNN6, except that the degenerate region is 5, 7, 8, and 9 codons long. In addition, oligonucleotides are made that code for various shapes of constrained random peptides by specifying sequences comprising three cysteine residues interspersed between 6-10 randomly specified amino acids.

20 6.3.2. IN VITRO SYNTHESIS AND ISOLATION OF POLYSOMES

An E. coli S30 extract is prepared from the B strain SL119 (Promega). Coupled transcription-translation reactions are performed by mixing the S30 extract with the S30 premix 25 (containing all 20 amino acids), the linear DNA template coding for peptides of random sequences (prepared as described in Section 6.3.1 above), and rifampicin at 20 μ g/ml. The reaction is initiated by the addition of 100 units of T7 RNA polymerase and continues at 37°C for 30 min. 30 The reaction is terminated by placing the reactions on ice and diluting them 4-fold with polysome buffer (20 mM Hepes-NaOH, pH 7.5, 10 mM MgCl₂, 1.5 μ g/ml chloramphenicol, 100 μg/ml acetylated bovine serum albumin, 1 mM dithiothreitol, 20 units/ml RNasin, and 0.1% Triton X-100). Polysomes are isolated from a 50 μ l reaction programmed with 0.5-1 μ g of linear DNA template specifying the synthesis of random constrained peptides. To isolate polysomes, the diluted S30

reaction mixtures are centrifuged at 288,000 X g for 30-40 min at 4°C. The pellets are suspended in polysome buffer and centrifuged a second time at 10,000 X g for 5 min to remove insoluble material.

5

6.3.3. AFFINITY SELECTION/SCREENING OF POLYSOMES

The isolated polysomes are incubated in microtiter wells coated with the target proteins. Microtiter wells are uniformly coated with 1-5 µg of 6-His tagged, or glutathione S-transferase fused, target proteins (see Section 6.2 hereinabove). Target proteins that are used include the oncoproteins ras and raf, KDR (the vascular endothelial growth factor [vEGF] receptor protein) and vEGF. The microtiter wells are coated with 1-5 μg of these target 15 proteins by incubation in PBS (phosphate-buffered saline; 10 mM sodium phosphate, pH 7.4, 140 mM NaCl, 2.7 mM KCl), for 1-5 hours at 37°C. The wells are then washed with PBS, and the unbound surfaces of the wells blocked by incubation with PBS containing 1% nonfat milk for 1 hr at 37°C. Following a wash 20 with polysome buffer, each well is incubated with polysomes isolated from a single 50 μ l reaction for 2-24 hr at 4°C. Each well is washed five times with polysome buffer and the associated mRNA is eluted with polysome buffer containing 20 mM EDTA

After affinity selection of the polysomes, the associated mRNAs are isolated, and treated with 5-10 units of DNase I (RNase-free; Ambion) for 15 min at 37°C after addition of MgCl₂ to 40 mM. The mRNA is phenol-extracted and ethanol-precipitated and dissolved in 20 μl of RNase-free water. A portion of the mRNA is used for cDNA preparation and subsequent amplification using 15 pmol each of primers RGSPCR [5'ATCAAGTTTGCCTTTACCAGCATTGTGGAGCGCGTTTTCATC3' (SEQ ID NO:5)], and SELEXF1

[5'ACTTCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCC3'

(SEQ ID NO:9)] and rTth Reverse Transcriptase RNA PCR kit

(Perkin Elmer Cetus). Specifically, the mRNA is reverse-

transcribed into cDNA in a 20 µl reaction containing 1 pg mRNA, 15 pmol of RGSPCR primer, 200 μM each of dGTP, dATP, dTTP, and dCTP, 1 mM MnCl2, 10 mM Tris-HCl, pH 8.3, 90 mM KCl, and 5 units of rTth DNA polymerase at 70°C for 15 min. 5 the next step, the cDNA is amplified by the addition of 2.5 mM MgCl,, 8% glycerol, 80 mM Tris-HCl, pH 8.3, 125 mM KCl, 0.95 mM EGTA, 0.6% Tween 20, and 15 pmol of the SELEXF1 primer. The reaction conditions that are employed are 2 min at 95°C for one cycle, 1 min at 95°C and 1 min at 60°C for 35 10 cycles, and 7 min at 60°C for one cycle. The amplified product is then gel-purified and quantitated by spectrophotometry at 260 nm. A portion of the amplified DNA is digested with NsiI and XbaI and the resulting 30 base pair fragment is directionally cloned into a monovalent phage 15 display vector. The DNAs inserted in the monovalent phage display vector are then sequenced to determine the identity of the peptides that were selectively retained by one cycle of affinity binding to the target protein. A second portion $(0.5-1 \mu g)$ of the amplified DNA is subjected to another cycle 20 of affinity selection, mRNA isolation, cDNA amplification, and cloning.

6.4. PHAGEMID SCREENING

Three different protocols for screening of a phagemid 25 library are presented in the subsections hereinbelow. These protocols, particularly the immobilization and binding steps, are readily adaptable to use for screening of different libraries, e.g., polysome libraries. Preferably, different methods are used in different rounds of screening.

30

6.4.1. PLATE PROTOCOL

In this example, a protocol is presented for screening a phagemid library, in which in the first round of screening, a 35 biotinylated target protein is immobilized (by the specific binding between biotin and streptavidin) on a streptavidin

coated plate. The immobilized target protein is then contacted with library members to select binders.

Reagents Used:

- 5 Purified target protein, microfuge tubes, Falcon 2059, Binding Buffer, Wash Buffer, Elute Buffer, phage display Library of >10¹¹ pfu/Screened Target, fresh overnight cultures of appropriate host cells, LB Agar plates with antibiotics as needed, biotinylating agent NHS-LC-Biotin (Pierce Cat.
- 10 #21335), streptavidin, 50 mM NaHCO, pH 8.5, 1 M Tris pH 9.1, M280 Sheep anti-mouse IgG coated Dynabeads (Dynal), phosphate buffered saline (PBS), Falcon 1008 petri dishes.

Wash Buffer = 1X PBS (Sigma Tablets), 1 mM MgCl₂, 1 mM CaCl₂, 15 0.05% Tween 20; (For one liter: 5 PBS tablets, 1 ml 1 M MgCl₂, 1 ml 1 M CaCl₂, 0.5ml Tween 20, nanopure H₂O to 1 liter).

Binding Buffer = Wash Buffer with 5 mg/ml bovine serum albumin (BSA).

20

Elute Buffer = 0.1 N HCl adjusted to pH 2.2 with glycine: 1 mg/ml BSA.

Procedure:

25 Protein Biotinylation:

- 1. Wash 50-100 μ g of target protein in 50 mM NaHCO, pH 8.5 in a Centricon (Amicon) of the appropriate molecular weight cut-off.
- 2. Bring the total volume to 100 μ l with 50 mM NaHCO, pH 30 8.5.
 - 3. Dissolve 1 mg of NHS-LC-Biotin in 1 ml $\rm H_2O$. Do not store this solution.
- 4. Immediately add 37 μ l of the NHS-LC-Biotin solution to the target protein and incubate for 1 hr at room temperature 35 (RT).

5. Remove the unreacted biotin by washing 2X PBS in a Centricon (Amicon) of the appropriate molecular weight cutoff. Store the biotinylated protein at 4°C.

5 Coating a 1008 Plate with Streptavidin:

- 6. The night before the binding experiment precoat a 1008 plate with streptavidin.
- 7. Add 10 μ g of streptavidin (1 mg/ml H₂O) per 1 ml of 50 mM NaHCO₃ pH 8.5.
- 10 8. Add 1 ml of this solution to each plate and place in a humidified chamber overnight at 4°C.

Prebinding: Blocking Non-Specific Sites:

- 9. To a streptavidin coated plate add 400 μ l of Binding
- 15 Buffer (BSA blocking) for one hour at room temperature.
 - 10. Rinse wells six times with Wash Buffer by slapping dry on a clean piece of labmat.

Binding; Specific Target/Phage Complexes Round 1:

- 20 11. Add 10 μ g of biotinylated target protein in 400 μ l of Binding Buffer to the well and incubate for 2 hr at 4°C.
 - 12. Add 4 µl of 10 mM biotin and swirl for 1 hr at 4°C.
 - 13. Wash as in step 10.
- 14. Add concentrated phage library (>10¹¹ pfu) in 400 μ l of 25 Binding Buffer and swirl overnight at 4°C.

Washing and Elution:

- 15. Slap out binding mixture and wash as in step 10.
- 16. To elute bound phage add 400 μ l of Elution Buffer and 30 rock at RT for 15 min.
 - 17. Transfer the elution solution to a sterile 1.5 ml tube which contains 75 μ l of 1 M Tris pH 9.1. Vortex briefly.

Amplification of Round 1 Eluted Phage:

35 18. Plate all of the eluted round 1 phage by adding 157 μ l of phage to 200 μ l of cells incubated overnight (previously checked free of contamination) in three aliquots. Incubate

25 min in a 37°C water bath and then spread onto LB agar/antibiotics plate containing 2% glucose.

- 19. Scrape plates with 5 ml of 2XYT (growth broth) / Antibiotics/Glucose and leave swirling for 30 min at RT.
- 5 20. Add the appropriate amount of 2XYT/Antibiotics/Glucose to bring the O.D. 600 down to 0.4 and then grow at 37°C at 250 rpm until the O.D. 600 reaches 0.8.
 - 21. Remove 5 ml and add to it 1.25 x 1010 M13 helper phage.
- 22. Shake 30 min at 150 rpm and then 30 min at 250 rpm at 10 37°C.
 - 23. Centrifuge 10 min at 3000 X g at RT.
 - 24. Resuspend cells in 5 ml 2XYT with no glucose. (This step removes glucose).
- 25. Centrifuge as in step 23 and resuspend in 5 ml 2XYT with 15 kanamycin and the appropriate antibiotics (no glucose). Spin 18 hr at 37°C and 250 rpm.
 - 26. Pellet cells at 10,000 X g and sterile filter the phage containing supernatant which is now ready for round 2 screening.
- 20 27. Titer the round 1 eluted phage stocks.

Binding; Specific Target/Phage Complexes Rounds 2-5:

- 6. Combine -1 μ g of biotinylated target protein with the eluted and titered round 1 phage (10° pfu) in 200 μ l of
- 25 Binding Buffer and rock 4 hr at 4°C.
 - 7. The night before the round 2 screening is started, prewash 200 μ l/target protein to be screened of sheep antimouse IgG magnetic beads (M280 IgG Dynabeads) with 2X 1 ml of Wash Buffer using the Dynal Magnet. Let the beads collect at
- 30 least 1 min before removing the buffer. Let the beads stand 15 sec to allow residual Binding Buffer to collect and remove with a P200 Pipetman.
- 8. Resuspend the washed beads in 200 μ l of Binding Buffer and add 100 μ l of mouse anti-biotin IgG (Jackson IRL). Rock 35 overnight at 4°C.
 - 10. Wash the unbound anti-biotin IgG from the Dynabeads by placing them on the Dyna magnet for at least 1 min and remove

all liquid as in Step 7. Remove the tube from the magnet and resuspend the beads in 1 ml of Wash Buffer, rock at 4° C for 30 min, and return to the magnet. Again let the beads pellet for 1 min; repeat this process 3 more times and resuspend the 5 beads in 400 μ l of Binding Buffer.

- 10a. The coated beads are now ready for use (100 μ l/round/target protein). The remainder can be stored for use for up to 2 weeks.
- 11. Add the 100 μl of anti-biotin coated Dynabeads (Step 10)
 10 to the protein/phage fraction (Step 9) bringing the total binding volume to 300 μl and rock for 2 hr at 4°C. Ensure that the beads mix thoroughly with the phage/protein solution.

15 Washing and Elution:

- 12. Place the binding reaction into the Dynal magnet and let sit for 1 min.
- 13. Remove the solution using a P1000 Pipetman and discard. Let the beads stand 15 sec to allow residual binding buffer
- 20 to collect and remove with a P200 Pipetman. Note serial dilution depends upon all residual liquid being removed (i.e., 5 μ l into 500 is 100X washing; 50 μ l into 500 is only 10X).
- 14. Remove the tube from the magnet and resuspend the beads 25 in 750 μ l of Wash Buffer and return to the magnet. Again let the beads pellet by waiting 1 min.
 - 15. Remove the Wash solution as in Step 7 and repeat this process several more times.
- 16. After the removal of the final wash, resuspend the beads 30 and transfer them to a fresh, labeled tube and wash once more.
 - 17. To elute bound phage, add 400 μ l of Elution Buffer, titrate and rock for 14 min at RT.
- 18. Place the tube on the magnet for one minute and transfer 35 the eluate to a sterile 1.5 ml tube which contains 75 μ l of 1 M Tris pH 9.1. Vortex briefly.

Amplification of Round 2-5 Eluted Phage:

15a. Plate 10 μ l and 100 μ l of round 2,3,4 eluates using 200 μ l of contamination free (previously tested) *E. coli* XL1Blue cells onto each plate containing

5 tetracycline/ampicillin/glucose and tetracycline/ampicillin and amplify as in Steps 17-25.

6.4.2. BIOTIN-ANTIBIOTIN IGG BEAD PROTOCOL

In this example, a protocol is presented for screening a 10 phagemid library, in which a biotinylated target protein is immobilized (by the specific binding between anti-biotin antibodies and biotin) on a magnetic bead containing anti-biotin antibodies on the bead surface. The immobilized target protein is then contacted with library members to 15 select binders.

Reagents Used:

M280 Sheep anti-Mouse IgG coated Dynabeads (Dynal)

- 20 Binding; Specific Target/Phage Complexes Round 1:
 - 6. Combine 10 μ g of biotinylated target protein with the phage library (>10¹⁰ pfu) in 400 μ l of Binding Buffer and rock overnight at 4°C.
- 7. That same night prewash 50 μ l sheep anti-mouse IgG 25 magnetic beads (M280 IgG Dynabeads) with 500 μ l of Binding Buffer twice using the Dynal Magnet. Let the beads collect at least 1 min before removing the buffer. Let the beads stand 15 sec to allow residual binding buffer to collect and remove with a P200 Pipetman.
- 30 8. Resuspend the washed beads in 100 μ l of Binding Buffer and add 33 μ l of mouse anti-biotin IgG (40 μ g, Jackson IRL). Rock overnight at 4°C.
 - 9. Remove unbound protein from the phage/protein reaction in Step 6 with a Microcon 100. Spin at 800 X g until
- 35 exclusion volume is met and wash twice with Wash Buffer (again at 800 X g). Collect phage/protein with a Pipetman and add an additional 50 μ l of Wash Buffer to the Microcon,

gently titrate and combine with first fraction to ensure maximal recovery.

- 10. Wash the unbound anti-biotin IgG from the Dynabeads by placing them on the Dyna magnet for at least 1 min and remove
- 5 all liquid as in Step 7. Remove the tube from the magnet and resuspend the beads in 750 μ l of Wash Buffer, rock at 4°C for 30 min, and return to the magnet. Again, let the beads pellet for 1 min; repeat this process 3 more times and resuspend the beads in 100 μ l of Binding Buffer.
- 10 11. Add the anti-biotin coated Dynabeads (Step 10) to the protein/phage fraction (Step 9), bring the total binding volume to 500 µl with Binding Buffer, and rock for 2 hr at RT. Ensure that the beads mix thoroughly with the phage/protein solution.

15

Washing and Elution:

- 12. Place the binding reaction into the Dynal magnet and let sit for 1 min.
- 13. Remove the solution using a P1000 Pipetman and discard.
- 20 Let the beads stand 15 sec to allow residual binding buffer to collect and remove with a P200 Pipetman. Note that serial dilution depends upon all residual liquid being removed (i.e., 5 μ l into 500 is 100% washing; 50 μ l into 500 is only 10%).
- 25 14. Remove the tube from the magnet and resuspend the beads in 750 μl of Wash Buffer and return to the magnet. Again let the beads pellet by waiting 1 min.
 - 15. Remove the wash solution as in Step 7 and repeat this process 3 more times.
- 30 16. After the removal of the fourth wash, resuspend the beads and transfer them to a fresh, labeled tube and wash
 - 17. To elute bound phage, add 400 μ l of Elution Buffer, titrate and rock for 14 min at RT.
- 35 18. Place the tube on the magnet for one minute and transfer the eluate to a sterile 1.5 ml tube which contains 75 μ l of 1 M Tris pH 9.1. Vortex briefly.

Amplification of Round 1 Eluted Phage:

17. Plate all of the eluted round 1 phage by adding 157 μ l of phage to 200 ml of cells incubated overnight (previously checked to be free of contamination) in three aliquots.

- 5 Incubate 25 min in a 37°C water bath and then spread onto LB agar/antibiotics plate containing 2% glucose. Place plates upright in 37°C incubator until dry and then invert and incubate overnight.
- 18. Scrape plates with 5 ml of 2XYT/Antibiotics/Glucose and 10 leave swirling for 30 min at RT.
- 19. Add the appropriate amount of 2XYT/Antibiotics/Glucose to bring the O.D. 600 down to 0.4 and then grow at 37°C at 250 rpm until the O.D. 600 reaches 0.8.
 - 20. Remove 5 ml and add to it 1.25 x 1010 M13 helper phage.
- 15 21. Shake 30 min at 150 rpm and then 30 min at 250 rpm at 37°C.
 - 22. Centrifuge 10 min at 3000 X g at RT.
 - 23. Resuspend cells in 5 ml 2XYT with no glucose. (This step removes glucose)
- 20 24. Centrifuge as in step 23 and resuspend in 5 ml 2XYT with kanamycin and the appropriate antibiotics (no glucose). Spin 18 hr at 37°C and 250 rpm.
 - 25. Pellet cells at 10,000 xg and sterile filter the phage-containing supernatant which is now ready for round 2
- 25 screening.

Binding; Specific Target/Phage Complexes Round 2, 3, & 4: 6a. Bind 1 μ g of target protein with 100 μ l of amplified phage from the previous round as before, overnight at 4°C.

- 30 7a. Prepare the IgG anti biotin/anti IgG beads as in Steps 7-10 using, however, only 20 μ l of sheep anti-mouse IgG and 13 μ l of anti-biotin IgG.
 - 8a. All other binding procedures are identical with Steps 6-11.

Washing and Elution:

9a. Place the binding reaction into the Dynal magnet and let sit for 1 min.

- 10a. Remove the solution and discard using a P1000 Pipetman.
- 5 Let the beads stand 30 sec to allow residual Binding Buffer to collect and remove with a P200 Pipetman.
 - 11a. Remove the tube from the magnet and resuspend the beads in 750 μl of Wash Buffer and return to the magnet. Again let the beads pellet by waiting 1 min.
- 10 12a. Remove the wash solution as in Step 11a and repeat this process 3 more times.
 - 13a. After the removal of the fourth wash, resuspend the beads and transfer them to a fresh, labeled tube and wash 4 more times.
- 15 14a. Elute and neutralize as in Step 15.

Amplification of Rounds 2, 3, & 4 Eluted Phage: 15a. Plate 10 μ l and 100 μ l of round 2,3,4 eluates and amplify as in Steps 17-25.

20

6.4.3. BIOTIN-STREPTAVIDIN, MAGNETIC BEAD PROTOCOLS

In this example, a protocol is presented for screening a phagemid library, in which a biotinylated target protein is immobilized (by the specific binding between biotin and streptavidin) on a streptavidin coated magnetic bead. The immobilized target protein is then contacted with library members to select binders.

30 Reagents Used:

Purified target protein, M280 streptavidin coated Dynabeads (Dynal)

Binding; Specific Target/Phage Complexes Round 1:

35 6. Combine 10 μ g of biotinylated target protein with the phage library (>10¹⁰ pfu) in 400 μ l of Binding Buffer and rock overnight at 4°C.

7. Remove unbound protein with a Microcon 100. Spin at 800 X g until exclusion volume is met, and wash twice with Wash Buffer (again at 800 X g). Collect phage/protein with a Pipetman and add an addition 50 μ l of Wash Buffer to the Microcon Cently titrate and combine with a

- 5 Microcon, gently titrate and combine with the first fraction to ensure maximal recovery.
 - 8. Prewash 50 μ l (per reaction) of streptavidin magnetic beads (M280 streptavidin Dynabeads) twice with 500 μ l of Washing Buffer using the Dynal magnet.
- 10 9. Add the prewashed Dynabeads to the protein/phage fraction (add Binding Buffer to a total of 500 μ l) and rock for 30 min. Ensure that the beads mix thoroughly with the phage/protein solution.

15 Washing and Elution:

- 10. Place the binding reaction into the Dynal magnet and let sit for 1 min.
- 11. Remove the solution using a P1000 Pipetman and discard. Let the beads stand 15 sec to allow residual Binding Buffer to
- 20 collect and remove with a P200 Pipetman. Note that serial dilution depends upon all residual liquid being removed (i.e., 5 μ l into 500 is 100X washing; 50 μ l into 500 is only 10X).
- 12. Remove the tube from the magnet and resuspend the beads in 750 μ l of Wash Buffer and return to the magnet. Again let 25 the beads pellet by waiting 1 min.
 - 13. Remove the wash solution as in step 11 and repeat this process 3 more times.
 - 14. After the removal of the fourth wash, resuspend the beads and transfer them to a fresh, labeled tube and wash once more.
- 30 15. To elute bound phage add 400 μl of Elution Buffer, titrate and rock for 14 min at RT.
 - 16. Place the tube on the magnet for one minute and transfer the eluate to a sterile 1.5 ml tube which contains 75 μ l of 1 M Tris pH 9.1. Vortex briefly.

Amplification of Round 1 Eluted Phage:

17. Plate all of the eluted round 1 phage by adding 157 μ l of phage to 200 μ l of overnight cells (previously checked to be free of contamination) in three aliquots. Incubate 25 min in

- 5 a 37°C water bath and then spread onto LB agar/antibiotics plate containing 2% glucose. Place plates upright in 37°C incubator until dry and then invert and incubate overnight.
 - 18. Scrape plates with 5 μ l of 2XYT/Antibiotics/Glucose and leave swirling for 30 min at RT.
- 10 19. Add the appropriate amount of 2XYT/Antibiotics/Glucose to bring the O.D. 600 down to 0.4 and then grow at 37°C at 250 rpm until the O.D. 600 reaches 0.8.
 - 20. Remove 5 ml and add to it 1.25 x 1010 M13 helper phage.
- 21. Shake 30 min at 150 rpm and then 30 min at 250 rpm at 15 37°C.
 - 22. Centrifuge 10 min at 3000 X g at RT.
 - 23. Resuspend cells in 5 μ l 2XYT with no glucose. (This step removes glucose).
- 24. Centrifuge as in step 22 and resuspend in 5 ml 2XYT with 20 kanamycin and the appropriate antibiotics (no glucose). Shake 18 hr at 37°C and 250 rpm.
 - 25. Pellet cells at 10,000 X g and sterile filter the phage containing supernatant which is now ready for round 2 screening.

- Binding; Specific Target/Phage Complexes Round 2, 3, & 4: 6a. Combine 1 μ g of biotinylated target protein with 100 μ l of the previous round's phage (>10° pfu) in 400 μ l of Binding Buffer and rock overnight at 4°C.
- 30 7a. Remove unbound protein with a Microcon 100. Spin at 800 X g until exclusion volume is met and wash twice with Wash Buffer (again at 800 X g). Collect phage/protein with a Pipetman and add an addition 50 μ l of Wash Buffer to the Microcon, gently titrate and combine with the first fraction 35 to ensure maximal recovery.

8a. Prewash 20 μ l (per reaction) of streptavidin magnetic beads (M280 streptavidin Dynabeads) twice with 500 μ l of Washing Buffer using the Dynal magnet.

9a. Add the prewashed Dynabeads to the protein/phage fraction 5 and rock for 30 min. Add Binding Buffer to a total of 500 μ l. Ensure that the beads mix thoroughly with the phage/protein solution.

Washing and Elution:

30

- 10 10a. Place the binding reaction into the Dynal magnet and let sit for 1 min.
 - 11a. Remove the solution and discard using a P1000 Pipetman. Let the beads stand 30 sec to allow residual Binding Buffer to collect and remove with a P200 Pipetman.
- 15 12a. Remove the tube from the magnet and resuspend the beads in 750 μl of Wash Buffer and return to the magnet. Again let the beads pellet by waiting 1 min.
 - 13a. Remove the wash solution as in Step 11a and repeat this process 3 more times.
- 20 14a. After the removal of the fourth wash resuspend the beads and transfer them to a fresh, labeled tube and wash 4 more times.
 - 15a. Elute and neutralize as in Step 15.
- 25 Amplification of Rounds 2, 3, & 4 Eluted Phage: 16a. Plate 10 μ l and 100 μ l of round 2,3,4 eluates and amplify as in Steps 17-25.

6.5. AFFINITY MEASUREMENTS OF PEPTIDE-TARGET PROTEIN INTERACTIONS

Once peptides that bind to a target protein have been identified, the affinities of these peptides to their respective targets are measured by measuring the dissociation constants (K_d) of each of these peptides to their respective targets. Oligonucleotides that encode the peptides are constructed so as to encode also an epitope tag fused to the peptide (for example, the myc epitope) that can be detected by

a commercially available antibody. These oligonucleotides are incubated with polysome extracts to produce the peptide tagged with the epitope. Binding of the target protein to the peptide is done in solution, and separation of the bound 5 peptide from the unbound peptide is done by immunoaffinity purification using an anti-target protein antibody. immunoaffinity purification is done by a modified ELISA (enzyme-linked immunosorbent assay) protocol, in which the target protein-peptide mixture is exposed to the anti-target 10 protein antibody immobilized on a solid support such as a nitrocellulose membrane, and the unbound peptide is then washed off. In this protocol, the concentration of the target protein is varied and then the amount of bound peptide is estimated by detecting the epitope tag on the peptide by use 15 of anti-epitope antibody. In this manner, the affinity of each peptide for its target protein can be determined.

6.6. REDOR MEASUREMENTS ON A CX,C PEPTIDE RESIN

This example demonstrates successful synthesis and 20 cyclization of a CX₆C peptide resin of greater than 95% purity and with a labeled glycine followed by successful REDOR distance measurements on the CX₆C peptide resin using the preferred REDOR methods of this invention. The labeled peptide used was

25 Cys-Asn-Thr-Leu-Lys-(15N-2-13C)Gly-Asp-Cys-Gly-mBHA resin, where a glycine linker attached the peptide of interest to the nBHA resin. (Cys-Asn-Thr-Leu-Lys-Gly-Asp-Cys-Gly = SEQ ID NO:10)

The peptide resin was synthesized by solid phase synthesis on p-MethylBenzhydrilamine (mBHA) resin using a 30 combination of Boc and Fmoc chemistry. MethylBenzhydrilamine resin (Subst. 0.36 meq/g) was purchased from Advanced Chem Tech (Louisville, KY). Fmoc(15N-2-13C)Gly was prepared from HCl, (15N-2-13C)Gly (Isotec Inc., Miamisburg, OH) and Fmoc-OSu. Boc-Gly, (Trt), Fmoc-Asp(OtBu), Fmoc-Lys(Boc), Fmoc-Leu,

35 Fmoc-Thr(OtBu), Fmoc-Asn and Boc-Cys(Acm) were purchased from Bachem (Torrance, CA). Reagent grade solvents were purchased from Fisher Scientific, Diisopropylcarbodiimide (DIC),

Trifluoroacetic acid (TFA) and Diisopropylethylamine (DIEA) were purchased from Chem Impex (Wooddale, IL). Nitrogen, HF were purchased from Air Products (San Diego, CA).

The first step 43 was the synthesis of

- 5 Boc-Cys(ACM) -Asn-Thr(OtBu) -Leu-Lys(Boc) -Gly-Asp(OtBu) Cys(Trt) -Gly-mBHA resin. 1.11g (0.40 meq) of mBHA resin were placed in a 150 ml reaction vessel (glass filter at the bottom) with Methylene Chloride (CH₂Cl₂) ["DCM"] and stirred 15 min with a gentle bubbling of Nitrogen in order to swell the
- 10 resin. The solvent was drained and the resin was neutralized with DIEA 5% in DCM (3X2 min). After washes with DCM, the resin was coupled 60 min with Boc-Gly (0.280 g-1.6 meq-4 fold excess-0.1M) and DIC (0.25 ml-1.6 meq-4 fold excess-0.1M) in DCM. Completion of the coupling was checked with the
- 15 Ninhydrin test. After washes, the resin was stirred 30 min in TFA 55% in DCM in order to remove the Boc protecting group. The resin was then neutralized with DIEA 5% in DCM and coupled with Fmoc-Cys(Trt)(0.937g-1.6 meq-4 fold excess-0.1M) and DIC (0.25 ml-1.6 meq-4 fold excess-0.1M) in DCM/DMF (50/50).
- 20 After washes the resin was stirred with Piperidine 20% in DMF (5 min and 20 min) in order to remove the Fmoc group. After washes, this same cycle was repeated with Fmoc-Asp(OtBu), Fmoc(15N-2-13C)Gly (2 fold excess only), Fmoc-Lys(Boc), Fmoc-Leu, Fmoc-Thr(OtBu), Fmoc-Asn and Boc-Cys(Acm). After the
- 25 last coupling, the Boc group was left on the peptide. The resin was washed thoroughly with DCM and dried under a nitrogen stream. Yield was 1.49g (Expected: ~1.7g).

The next step 44 was cyclization of the Boc-Cys-Asn-Thr(OtBu)-Leu-Lys(Boc)-Gly-Asp(OtBu)-Cys-Gly-mBHA 30 resin. 600 mg of protected peptide resin were sealed in a polypropylene mesh packet. The bag was shaken in a mixture of solvent (DCM/Methanol/Water-640/280/47) in order to swell the resin. The bag was then shaken 20 min in 100 ml of a solution of iodine in the same mixture of solvent (0.4 mg I₂/ml solvent 35 mixture). This operation was performed 4 times. No decoloration was observed after the third time. The resin was

then thoroughly washed with DCM, DMF, DCM, and methanol successively.

The last step 45 was side-chain deprotection of the Cys-Asn-Thr-Leu-Lys-Gly-Asp-Cys-Gly-mBHA resin. After 5 cyclization the resin in the polypropylene bag was reacted 1.5 hour with 100 ml of a mixture TFA/p-Cresol-Water (95/2.5/2.5). After washes with DCM and Methanol, the resin was dried 48 hours under vacuum. Yield was 560 mg.

The resulting peptide resin was analyzed for its purity

10 and the presence of the disulfide bridge. 40 mg of resin were
sealed in a propylene mesh packet and treated with HF at 0 C
for 1 hour in presence of anisole (HF/Anisole: 90/10). The
scavenger and by-products were extracted from the resin with
cold ethyl ether. The peptide was extracted with 10% Acetic

15 Acid and lyophilized 36 hours. The dry isolated peptide was
characterized by PDMS (mass spectrography) and HPLC (high
performance liquid chromatography). This analysis
demonstrated that greater than 95% of the product peptide was
of the correct amino acid composition, having a disulfide loop

20 and without inter-molecular disulfide dimers.

REDOR measurements were made on the peptide resin prepared by this method, and as a control, also on dried (15N-2-11C) labeled glycine. The preferred REDOR methods and parameters, as previously detailed, were used. Fig. 6
25 illustrates the 15N resonance spectral signals obtained. Signal 70 is the signal produced by dried glycine after no rotor periods. Signals 71, 72, 73 are glycine signals after 2, 4, and 8 rotor periods, respectively. Signals 74, 75, 76, and 77 are the peptide resin signals after 0, 2, 4, and 8 rotor periods, respectively.

Fig. 7 illustrates the data analysis. As in Fig. 5, axis 81 is the ΔS/S axis, and axis 82 is the λ axis. The variables are as used in equation 5. Graph 83 is defined by equation 5, and is the initial rising part of the full curve shown in Fig. 35 5. Data points 84, 85, 86, and 87 are best fits of the data for 0, 2, 4, and 8 rotor periods, respectively. At these points, the circles represent the glycine values and the

squares the peptide resin values. These values correspond to a C-N distance in glycine and the peptide of 1.55 Å (and a D_{CK} of 800 Hz). Repeated measurements gave a C-N distance of 1.50 Å (and a D_{CK} of 875 Hz). The accepted distance in glycine is 1.48 Å. The above procedure was repeated for (15N-1-13C) labeled glycine in Cys-Asn-Thr-Leu-Lys-(15N-1-13C)Gly-Asp-Cys-Gly-mBHA resin, and the measured C-N distance of 2.50 Å is in excellent agreement with the predicted value of 2.46 Å.

Thus REDOR accuracy to better that 0.1 Å is demonstrated. Also demonstrated is the peptide resin as an appropriate substrate for NMR measurements. Inter-molecular dipole-dipole interactions between adjacent peptides did not interfere. Also the overlap of the distances measured in free glycine and 15 in glycine incorporated in the peptide demonstrated that the peptide was held sufficiently rigidly by the resin that any remaining peptide motions did not interfere with the NMR measurements.

20 7. SPECIFIC EMBODIMENTS, CITATION OF REFERENCES

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from 25 the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

30

WHAT IS CLAIMED IS:

 A method of determining a consensus pharmacophore structure comprising the steps of:

- 5 (a) identifying from one or more diversity libraries a plurality of compounds that bind to a target molecule,
 - (b) measuring one or more distances in one or more of the compounds, and
- (c) determining a consensus pharmacophore structure for the compounds.
 - The method of claim 1 wherein said compounds are peptides, peptide derivatives, or peptide analogs.

15

- 3. The method of claim 2 wherein said compounds are peptides containing one or more cystines.
- 4. The method of claim 3 wherein the peptides comprise the sequence CX₄C (SEQ ID NO:1).
- The method of claim 1 further comprising a step of selecting a plurality of candidate pharmacophores based on rules of chemical homology, the selected plurality of candidate pharmacophores being used in step (c) to determine the consensus pharmacophore structure.
- The method of claim 5 wherein the rules of homology determine that two candidate pharmacophores are homologous if they have chemically similar side chains.
 - 7. The method of claim 1 which further comprises after said identifying step, a screening step involving a genetic selection technique.

35

8. The method of claim 1 wherein the step of measuring distance comprises making solid phase nuclear magnetic

resonance measurements on selected nuclei in a nuclear magnetic resonance spectrometer upon a sample comprising one of the compounds.

5 9. The method of claim 8 wherein the step of measuring distances further comprises making rotational echo double resonance nuclear magnetic resonance measurements of internuclear dipole-dipole interaction strength between selected nuclei in the compound in the sample.

10

- 10. The method of claim 8 wherein the sample further comprises a substrate having a surface to which the compound is attached.
- 15 11. The method of claim 8 wherein the sample is cooled below room temperature.
 - 12. The method of claim 8 wherein the compound is bound to the target molecule.

20

25

- 13. The method of claim 10 wherein a plurality of the compound is attached to the surface at a surface density such that the inter-nuclear dipole-dipole interactions between different molecules is less than 10% of the inter-nuclear dipole-dipole interaction within one molecule.
- 14. The method of claim 10 wherein the substrate has pores of sufficient size to permit the target to diffuse and bind to the compound in the sample.
 - 15. The method of claim 9 wherein rotational echo double resonance nuclear magnetic resonance measurements can be made on the compound bound to the target or hydrated or in a dry nitrogen atmosphere.

16. The method of claim 10 wherein the compound is a peptide, and a plurality of the peptide is attached to the substrate surface, which has a purity of the peptide of at least 95% and wherein the surface density of the peptide is no more than one peptide per 100 Å² of substrate surface.

17. The method of claim 10 wherein the substrate is selected from the group consisting of p-MethylBenzhydrilamine resin, divinylbenzyl polystyrene resin, and glass beads.

- 18. The method of claim 8 wherein the selected nuclei are selected from the group consisting of ¹³C, ¹⁵N, ¹⁹F, and ³¹P.
- 15 19. The method of claim 9 wherein the nuclear magnetic resonance spectrometer comprises magnetic excitation means, a sample rotor, and free induction decay observing means, and the step of making rotational echo double resonance nuclear magnetic resonance measurements further comprises the steps of:
 - (a) spinning the sample in the sample rotor,
 - (b) initially exciting magnetically the selected nuclei to be observed,
- (c) providing subsequently one π spin flip magnetic excitation during each rotor period to each of the selected nuclei, the pulses to the different nuclei having fixed phase delays,
 - (d) observing the free induction decay signal as a function of the number of rotor periods; and
- (e) finding the dipole-dipole strength between the selected nuclei, whereby the internuclear distance between the selected nuclei can be obtained.
- 20. The method of claim 1 wherein the step of measuring distances comprises making liquid phase nuclear magnetic resonance measurements.

21. A method of determining a consensus pharmacophore structure comprising the st ps of:

- (a) identifying from one or more diversity libraries a plurality of compounds that bind to a target molecule,
- (b) determining a consensus pharmacophore structure for the compounds.
- 22. A method of determining a consensus pharmacophorestructure comprising the steps of:
 - (a) measuring one or more distances in one or more compounds that bind to a target molecule, and
 - (b) determining a consensus pharmacophore structure for the compounds.

15

20

- 23. The method of claim 21 or 22 further comprising a step of selecting a plurality of candidate pharmacophores based on rules of chemical homology, the selected plurality of candidate pharmacophores being used in step (b) to determine the consensus pharmacophore structure.
- 24. The method of claim 23 wherein the compounds have limited conformational degrees of freedom at the temperature of interest, and wherein the step of determining a consensus pharmacophore structure for each compound further comprises, performing a consensus configurational bias Monte Carlo method, said Monte Carlo method comprising the steps of:
- (a) generating a proposed structure for a compound identified from said one or more diversity libraries by making conformational alterations consistent with the conformational degrees of freedom, the alterations being made to a representation of the compound's current chemical and conformational structure to generate a proposed representation, the proposed structure being generated with a bias

toward more acceptable configurations of lower energy, whereby the method is made more effici nt,

- (b) accepting and storing the proposed structure according to a probability depending on an energy determined for the proposed structure, and
- (c) repeating these steps until sufficient structures have been stored for each compound to permit statistically significant determination of an equilibrium structure for each compound.

10

15

- 25. A method of determining one or more lead compounds for use as a drug that binds to a target molecule comprising the steps of:
- (a) identifying from one or more diversity libraries a plurality of compounds that bind to a target molecule;
 - (b) determining a consensus pharmacophore structure for the compounds; and
- (c) determining one or more lead compounds for use as a drug which share a pharmacophore specification with the determined consensus pharmacophore structure.
- 26. A method of determining one or more lead compounds for use as a drug that binds to a target molecule comprising the steps of:
 - (a) measuring one or more distances in one or more compounds that bind to a target molecule;
 - (b) determining a consensus pharmacophore structure for the compounds; and
- (c) determining one or more lead compounds for use as a drug which share a pharmacophore specification with the determined consensus pharmacophore structure.
- 27. The method according to claim 25 or 26 wherein said step of determining one or more lead compounds comprises modifying a compound identified as binding to the target molecule, said modification being done outside of the

pharmacophore structure, to render the compound more attractive for use as a drug.

- 28. The method of claim 5 wherein the compounds have limited conformational degrees of freedom at a temperature of interest, and wherein the step of determining a consensus pharmacophore structure for the compounds further comprises performing a consensus configurational bias Monte Carlo method, said Monte Carlo method comprising the steps of:
 - (a) generating a proposed structure for a compound identified from said one or more diversity libraries by making conformational alterations consistent with the conformational degrees of freedom, the alterations being made to a representation of the compound's current chemical and conformational structure to generate a proposed representation, the proposed structure being generated with a bias toward more acceptable configurations of lower energy,

15

- (b) accepting and storing the proposed structure according to a probability depending on an energy determined for the proposed structure, and
- (c) repeating these steps until sufficient structures

 have been stored for each compound to permit

 statistically significant determination of an

 equilibrium structure for each compound.
- 29. The method of claim 28 wherein the limited conformational degrees of freedom comprise torsional rotations about mutual bonds between otherwise rigid subunits of the compound, each rigid unit's representation comprising its interconnections and atomic composition, each atom's representation comprising its type and position, the torsional rotations respecting any conformational constraints present.

30. The method of claim 28 wherein the compound is a peptide, peptide derivative, or peptide analog.

- 31. The method of claim 28 wherein the conformational alterations comprise constrained, concerted torsional rotations or removal of a side chain and regrowth of the side chain with a new torsional conformation.
- 32. The method of claim 31 wherein the constrained, concerted torsional rotations are constrained so that no more than four rigid units are spatially displaced.
- The method of claim 28 wherein determining the energy for the proposed structure of one compound comprises
 including one or more constraint terms which represent knowledge of measured structure for the compound.
- 34. The method of claim 33 wherein the constraint terms comprise a weighted sum of squares of differences of the actual and measured structures.
- 35. The method of claim 28 wherein the energy is determined for the proposed structure of one compound by a method comprising including consensus terms which represent knowledge that the identified compounds all bind to the same target, the compounds being otherwise treated independently by the method.
- 36. The method of claim 35 wherein the consensus terms are a weighted sum of squares of differences in the atomic positions of a candidate pharmacophore from the average values of these positions in all the compounds.
- 37. The method of claim 35 wherein the step of determining the consensus pharmacophore structure comprises determining from the plurality of selected candidate pharmacophores a candidate pharmacophore for which the

consensus terms are relatively small compared to the total energy.

The method of claim 35 wherein the step of determining the consensus pharmacophore structure comprises determining from the plurality of selected candidate pharmacophores a candidate pharmacophore for which the consensus terms are minimum compared to other selected regions.

10

30

- 39. The method of claim 28 wherein the equilibrium structure is determined by a method comprising averaging selected generated and accepted structures for each compound.
- 15 40. The method of claim 39 wherein the averaging of structures comprises clustering selected generated and accepted structures into sets of similar structures and averaging these sets for each member.
- 20 41. A method of identifying a compound that binds to a target molecule comprising the following steps in the order stated:
 - (a) contacting compounds of a phage display or polysomebased diversity library with a target molecule;
- 25 (b) identifying one or more compounds in the library that bind to the target molecule;
 - first fusion protein comprising an identified compound, with a second fusion protein comprising the target molecule or a binding portion thereof, in which binding of the first fusion protein to the second fusion protein results in an increase in activity or activation of a transcriptional promoter or an origin of replication; and

(c) contacting one or more first fusion proteins, each

35 (d) identifying one or more of the compounds that when present in said first fusion protein result in said increase in activity or activation.

42. A method of making solid state nuclear magnetic resonance measurements comprising measuring internuclear dipoledipole interaction strengths between selected nuclei in a compound, said compound being attached to the surface of a substrate.

43. The method of claim 42 which further comprises before said measuring step the step of synthesizing a plurality of said compound on the surface of the substrate.

10

5

- 44. The method of claim 43 wherein said plurality of the compound is at least 95% pure.
- 45. The method of claim 42 wherein a plurality of said compound is attached to the substrate surface, with at least 10 Å spacing between molecules of the compound.
- 46. The method of claim 42 wherein the substrate has pores of sufficient size to permit a molecule to diffuse and bind to the compound.
- 47. The method of claim 42 wherein the substrate has a surface density of the compound such that the internuclear dipole-dipole interactions between different molecules of the compound is less than 10% of the internuclear dipole-dipole interaction within one molecule of the compound.
- 48. The method of claim 42 wherein the compound is a peptide, peptide derivative, or peptide analog.
 - 49. The method of claim 42 wherein the substrate is selected from the group consisting of p-MethylBenzhydrilamine resin, divinylbenzyl polystyrene resin, and a glass bead.

35

50. The method of claim 42 wherein said measuring step comprises using a nuclear magnetic resonance

spectrometer, said spectrometer comprising magnetic excitation means, a sample rotor, and free induction decay observing means; and said measurement of internuclear dipole-dipole interaction is done by a method comprising the steps of:

(a) spinning the sample in the sample rotor;

5

10

- (b) initially exciting magnetically the selected nuclei to be observed;
- (c) providing subsequently one or more π spin flip magnetic excitations during each rotor period to one or both of the selected nuclei, wherein pulses to the different nuclei have fixed phase delays:
- (d) observing a free induction decay signal as a function of the number of rotor periods; and
- (e) determining the dipole-dipole strength between the selected nuclei, whereby the internuclear distance between the selected nuclei can be obtained.
- 51. A method of configurational bias Monte Carlo

 20 determination of the structure of a compound having
 limited conformational degrees of freedom at a
 temperature of interest, the method comprising the steps
 of:
 - (a) generating a proposed structure for the compound by making conformational alterations consistent with the conformational degrees of freedom, the alterations being made to a representation of the compound's current chemical and conformational structure to generate a proposed representation;
- 30 (b) accepting and storing the proposed structure according to a probability depending on an energy determined for the proposed structure; and
- (c) repeating these steps until sufficient structures
 have been stored to permit statistically significant
 determination of an equilibrium structure.

5

20

52. The method of claim 51 wherein the conformational degrees of freedom comprise torsional rotations about mutual bonds between otherwise rigid subunits of the compound, each rigid unit's representation comprising its interconnections and atomic composition, each atom's representation comprising its type and position, the torsional rotations respecting any conformational

- 10 53. The method of claim 51 wherein the compound is a peptide, peptide derivative, or peptide analog.
- 54. The method of claim 51 wherein the conformational alterations comprise constrained, concerted torsional rotations.

constraints present.

55. The method of claim 54 wherein the constrained, concerted torsional rotations are constrained so that no more than four rigid units are spatially displaced.

56. The method of claim 51 wherein the conformational alterations comprise removal of a side chain and regrowth of the side chain with a new torsional conformation.

- 25 57. The method of claim 51 wherein the proposed structures are generated with a bias toward more acceptable configurations of lower energy.
- 58. The method of claim 51 wherein the energy is determined for the proposed structure by a method comprising including constraint terms which represent knowledge of measured structure for the compound.
- 59. The method of claim 58 wherein the constraint terms35 comprise a weighted sum of squares of differences of the actual and measured structures.

5

60. The method of claim 51 applied to a plurality of compounds of limited conformational degrees of freedom all of which bind to the same target molecule wherein the method further comprises a step of selecting a plurality of candidate pharmacophores based on rules of chemical homology.

- 61. The method of claim 60 wherein the energy is determined for the proposed structure of one of the plurality of compounds by a method comprising including consensus terms which represent knowledge that the compounds all bind to the same target molecule.
- 62. The method of claim 61 wherein the consensus terms are a
 weighted sum of squares of differences in the atomic
 positions of a candidate pharmacophore of said one of the
 plurality of compounds from the average values of these
 positions in all the compounds.
- 20 63. The method of claim 61 which further comprises a step of determining a consensus pharmacophore structure by determining from the plurality of selected candidate pharmacophores that candidate pharmacophore for which the consensus terms are minimum compared to other candidate pharmacophores.
- 64. The method of claim 60 which further comprises a step of determining a consensus pharmacophore structure by determining from the plurality of selected candidate pharmacophores that candidate pharmacophore for which the consensus terms are relatively small compared to the total energy.
- 65. The method of claim 63 or 64 which further comprises a step of determining one or more lead compounds for use as a drug which share a pharmacophore specification with the determin d consensus pharmacophore structure.

66. The method of claim 51 wherein the equilibrium structure is determined by a method comprising averaging selected generated and accepted structures.

- 5 67. The method of claim 66 wherein the averaging of structures comprises clustering selected generated and accepted structures into sets of similar structures and averaging these sets.
- 10 68. An apparatus for configurational bias Monte Carlo determination of the structure of a compound having limited conformational degrees of freedom at a temperature of interest, the apparatus comprising:
 - (a) memory means for storing

15

20

- (i) data structures representing the compound's chemical and conformational structure consistently with the compound's degrees of freedom,
 - (ii) similar data structures representing the compound's proposed structure and prior structures, and
 - (iii) parameters representing atomic interactions, and
- (b) processor means for executing programs for
- 25 (i) generating a proposed structure by making conformational alterations consistent with the conformational degrees of freedom and with a bias toward more acceptable configurations of lower energy,
 - (ii) accepting and storing the proposed structure according to a probability depending on an energy determined for the proposed structure, and
- (iii) repeating these steps until sufficient
 structures have been stored to permit
 statistically significant determination of an
 equilibrium structure.

69. The apparatus of claim 68 wherein the conformational degr es of fr edom comprise torsional rotations about mutual bonds between otherwise rigid subunits of the compound, each rigid unit's representation comprising its interconnections and atomic composition, each atom's representation comprising its type and position, the torsional rotations respecting any conformational constraints present.

- 10 70. The apparatus of claim 68 wherein the compound is a peptide, peptide derivative, or peptide analog.
- 71. The apparatus of claim 68 wherein the memory, processor, and control means are configured from a workstation type digital computer comprising RAM memory, disk memory, processor, and input and display devices.
- 72. The apparatus of claim 68 wherein the conformational alterations made by the processor means further comprise constrained, concerted torsional rotations or removal of a side chain and regrowth of the side chain with a new torsional conformation.
- 73. The apparatus of claim 72 wherein the constrained,
 25 concerted torsional rotations are constrained so that no
 more than four rigid units are spatially displaced.
- 74. The apparatus of claim 68 wherein the processor means determines an energy for the proposed structure by a method comprising including constraint terms which represent knowledge of measured structure for the compound.
- 75. The apparatus of claim 74 wherein the constraint terms

 comprise a weighted sum of squares of differences of the actual and measured structures.

76. The apparatus of claim 68 applied to a plurality of compounds of limited conformational degrees of freedom all of which bind to the same target molecule, and wherein the processor means further comprises programs for selecting a plurality of candidate pharmacophores based on rules of chemical homology.

- 77. The apparatus of claim 76 wherein the processor means determines an energy for the proposed structure of any one compound by a method comprising including consensus terms which represent knowledge that the compounds all bind to the same target molecule.
- 78. The apparatus of claim 77 wherein the consensus terms are
 a weighted sum of squares of differences in the atomic
 positions of the candidate pharmacophore of said one
 compound from the average values of these positions in
 all the compounds.
- 20 79. The apparatus of claim 77 wherein the processor means further comprises programs for determining a consensus pharmacophore structure by determining from the plurality of selected candidate pharmacophores a candidate pharmacophore for which the consensus terms are minimum compared to other candidate pharmacophores.
- 80. The apparatus of claim 77 wherein the processor means further comprises programs for determining a consensus pharmacophore structure by determining from the plurality of selected candidate pharmacophores a candidate pharmacophore for which the consensus terms are relatively small compared to the total energy.
- 81. The apparatus of claim 79 or 80 wherein the processor35 means further comprises programs for determining one or more lead compounds for use as a drug that share a

pharmacophore specification with the consensus pharmacophore structure.

- 82. The apparatus of claim 68 wherein the processor means determines an equilibrium structure by a method comprising averaging selected generated and accepted structures.
- 83. The apparatus of claim 82 wherein the averaging of

 10 structures further comprises clustering selected

 generated and accepted structures into sets of similar

 structures and averaging these sets.
- 84. In a digital computer, apparatus for configurational bias

 Monte Carlo determination of the structure of at least
 one compound having limited conformational degrees of
 freedom at a temperature of interest, said apparatus
 comprising:
- (a) first memory means for storing data structures representing the compound's chemical and conformational structure consistently with the compound's degrees of freedom,

- (b) second memory means for storing similar data structures representing the compound's proposed structure,
- (c) third memory means for storing similar data structures representing the compound's prior structures,
- (d) first processor means for generating a proposed structure by making conformational alterations consistent with the conformational degrees of freedom and with a bias toward conformations of lower energy,
- (e) second processor means for accepting and storing the proposed structure according to a probability depending on an energy determined for the proposed structure, and

(f) third processor means for controlling and repeating the generation and acceptance until sufficient structures have been stored to permit statistically significant determination of an equilibrium structure.

- 85. The digital computer apparatus of claim 84 wherein the conformational degrees of freedom comprise torsional rotations about mutual bonds between otherwise rigid
- subunits of the compound, each rigid unit's representation comprising its interconnections and atomic composition, each atom's representation comprising its type and position, the torsional rotations respecting any conformational constraints present.

15

5

- 86. The digital computer apparatus of claim 84 wherein the compound is a peptide, peptide derivative, or peptide analog.
- 20 87. The digital computer apparatus of claim 84 wherein the digital computer is a workstation type digital computer comprising RAM memory, disk memory, processor, and input and display devices.
- 25 88. The digital computer apparatus of claim 84 wherein the conformational alterations generated by the first processor means comprise constrained, concerted torsional rotations or removal of a side chain and regrowth of the side chain with a new torsional conformation.

30

89. The digital computer apparatus of claim 88 wherein the constrained, concerted torsional rotations are constrained so that no more than four rigid units are spatially displaced.

35

90. The digital computer apparatus of claim 84 wherein the second processor means det rmines an energy for the

proposed structure by a m thod comprising including constraint terms which represent knowledge of measured structure for the compound.

- 5 91. The digital computer apparatus of claim 90 wherein the constraint terms comprise a weighted sum of squares of differences of the actual and measured structures.
- 92. The digital computer apparatus of claim 84 in which said at least one compound is a plurality of compounds of limited conformational degrees of freedom all of which bind to the same target and wherein data are stored in said first memory means representing the chemical and conformational structure of said plurality of compounds and wherein the apparatus further comprises additional processor means for selecting a plurality of candidate pharmacophores based on rules of chemical homology.
- 93. The digital computer apparatus of claim 92 wherein the second processor means determines an energy for the proposed structure of one of said plurality of compounds by a method comprising including consensus terms which represent knowledge that the compounds all bind to the same target molecule.

25

30

- 94. The digital computer apparatus of claim 92 wherein the consensus terms are a weighted sum of squares of differences in the atomic positions of a candidate pharmacophore of said one of the plurality of compounds from the average values of these positions in all the compounds.
- 95. The digital computer apparatus of claim 93 wherein the apparatus further comprises processor means for determining a consensus pharmacophore structure by determining from the plurality of selected candidate pharmacophores a candidate pharmacophore for which the

consensus terms are relatively small compared to the total energy.

- 96. The digital computer apparatus of claim 93 wherein the apparatus further comprises processor means for determining a consensus pharmacophore structure by determining from the plurality of selected candidate pharmacophores a candidate pharmacophore for which the consensus terms are minimum compared to other candidate pharmacophores.
- 97. The digital computer apparatus of claims 95 or 96 wherein the apparatus further comprises processor means for determining one or more lead compounds for use as a drug that share a pharmacophore specification with the consensus pharmacophore structure.
- 98. The digital computer apparatus of claim 84 wherein the third processor means determines an equilibrium structure by a method comprising averaging selected generated and accepted structures.
- 99. The digital computer apparatus of claim 98 wherein the averaging of structures comprises clustering selected
 25 generated and accepted structures into sets of similar structures and averaging these sets.
- 100. In a digital computer, apparatus for configurational bias Monte Carlo determination of the structure of a plurality of compounds having limited conformational degrees of freedom, each compound having a backbone and side chains, said apparatus comprising:
 - (a) first memory means for storing data structures representing each compound's chemical and conformational structure consistently with that compound's degrees of freedom and constraints,

35

5

25

30

35

(b) second memory means for storing similar data structures representing a proposed structure for one or more of the compounds,

- (c) third memory means for storing similar data structures representing prior structures of the plurality of compounds,
- first processor means for generating a proposed (d) structure of a randomly selected compound by making conformational alterations consistent with the conformational degrees of freedom, the 10 conformational alterations being randomly distributed between alterations that alter the structure of a randomly selected side chain of the selected compound and alterations that alter the structure of a randomly selected region of the 15 backbone of the selected compound, the proposed structure being stored in the second memory means, the proposed structure being generated with a bias toward more acceptable structures of lower energy, whereby the method is made more efficient. 20
 - (e) second processor means for accepting a proposed structure according to a probability depending on an energy determined for the proposed structure, the energy including terms representing physical interactions and terms representing heuristic information about the compound's structure, the heuristic information comprising knowledge about measured distances in one or more compounds of said plurality and about the plurality of the compounds binding to a same target molecule,
 - (f) third processor means for controlling and repeating these steps until sufficient structures have been generated and accepted to permit statistically significant determination of an equilibrium structure.

5

20

25

101. The digital computer of claim 100 wherein the conformational degrees of freedom comprise torsional rotations about mutual bonds between otherwis rigid subunits of the compound, each rigid unit's representation comprising its interconnections and atomic composition, each atom's representation comprising its type and position, the torsional rotations respecting any

10 102. The digital computer of claim 100 wherein the compound is a peptide, peptide derivative, or peptide analog.

conformational constraints present.

- 103. A method of configurational bias Monte Carlo determination of the structure of a compound selected from the group consisting of a peptide, peptide derivative, and peptide analog, the method comprising the steps of:
 - (a) representing the conformation of the compound by interconnected rigid units capable of torsional rotation about common bonds, each rigid unit's representation comprising its interconnections and atomic composition, each atom's representation comprising its type and position,
 - (b) generating a proposed structure by making conformational alterations consistent with the compound's structure,
 - (c) accepting a proposed structure according to a probability depending on an energy determined for the proposed structure, and
- (d) repeating these steps until sufficient structures have been generated and accepted to permit statistically significant determination of an equilibrium structure.
- 35 104. An apparatus for configurational bias Monte Carlo determination of the structure of a compound selected

from the group consisting of a peptide, peptide derivative, and peptide analog, the apparatus comprising:

(a) memory means for storing

5

10

15

20

25

- (i) data structures representing the compound's conformation as interconnected rigid units capable of torsional rotation about common bonds, each rigid unit's representation comprising its interconnections and atomic composition, each atom's representation comprising its type and position,
- (ii) similar data structures representing the compound's proposed structure and prior structures, and
- (iii) parameters representing atomic interactions,
 and
- (b) processor means for executing programs for
 - (i) generating a proposed structure by making conformational alterations consistent with the compound's structure and with a bias toward more acceptable configurations of lower energy,
 - (ii) accepting a proposed structure according to a probability depending on an energy determined for the proposed structure, and
 - (iii) repeating these steps until sufficient structures have been generated and accepted to permit statistically significant determination of an equilibrium structure.
- 105. In a digital computer, apparatus for configurational bias

 Monte Carlo determination of the structure of a compound selected from the group consisting of a peptide, peptide derivative, and peptide analog, said apparatus comprising:
- (a) first memory means for storing data structures representing the compound's structure as interconnected rigid units capable of torsional rotation about common bonds, each rigid unit's

- 338 -

5

20

representation comprising its interconnections and atomic composition, each atom's representation comprising its type and position,

- (b) second memory means for storing similar data structures representing the compound's proposed structure,
- (c) third memory means for storing similar data structures representing the compound's prior structures.
- (d) first processor means for generating a proposed structure by making conformational alterations consistent with the compound's structure and constraints and with a bias toward conformations of lower energy,
- (e) second processor means for accepting a proposed structure according to a probability depending on an energy determined for the proposed structure, and
 - (f) third processor means for controlling and repeating these steps until sufficient structures have been generated and accepted to permit statistically significant determination of an equilibrium structure.
- Monte Carlo determination of the structure of a plurality of compounds selected from the group consisting of peptides, peptide derivatives, and peptide analogs, each compound having a backbone and side chains, said apparatus comprising:
- (a) first memory means for storing data structures representing each compound's structure as interconnected rigid units capable of torsional rotation about common bonds, each rigid unit's representation comprising its interconnections and atomic composition, each atom's representation comprising its type and position,

5

10

15

(b) second memory means for storing similar data structures representing a proposed structure for one or more of the compounds,

- (c) third memory means for storing similar data structures representing prior structures of the plurality of the compounds,
- (d) first processor means for generating a proposed structure of a randomly selected compound by making conformational alterations consistent with the compound's structure, the conformational alterations being randomly distributed between alterations that alter the structure of a randomly selected side chain of the selected compound and alterations that alter the structure of a randomly selected region of the backbone of the selected compound, the proposed structure being stored in the second memory means, the proposed structure being generated with a bias toward more acceptable structures of lower energy,
- structure according to a probability depending on an energy determined for the proposed structure, the energy including terms representing physical interactions and terms representing heuristic information about the compound's structure, the heuristic information comprising knowledge about measured distances in one or more compounds of said plurality and about the plurality of the compounds binding to a same target molecule,
- (f) third processor means for controlling and repeating
 these steps until sufficient structures have been
 generated and accepted to permit statistically
 significant determination of an equilibrium
 structure.
- 35 107. The method of claim 42 wherein the nuclear magnetic resonance is rotational echo double resonance.

108. The method of claim 1 wherein the diversity libraries are structurally constrained organic diversity libraries.

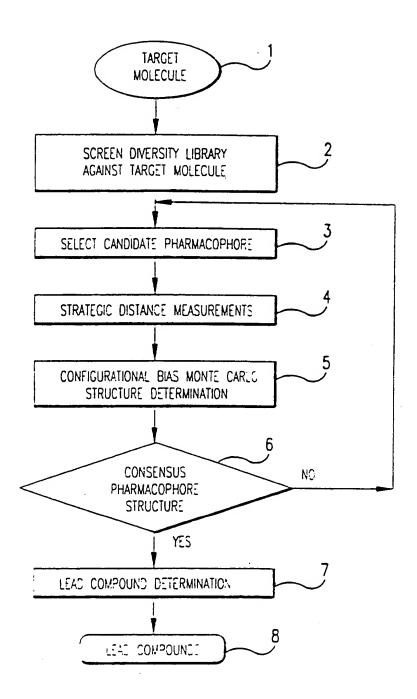


FIG.1

SUBSTITUTE SHEET (RULE 26)

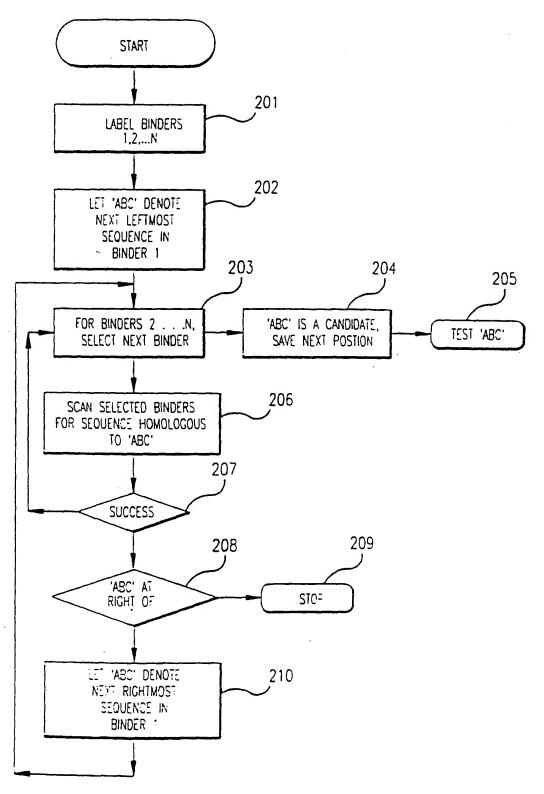


FIG.2A

SUBSTITUTE SHEET (RULE 26)

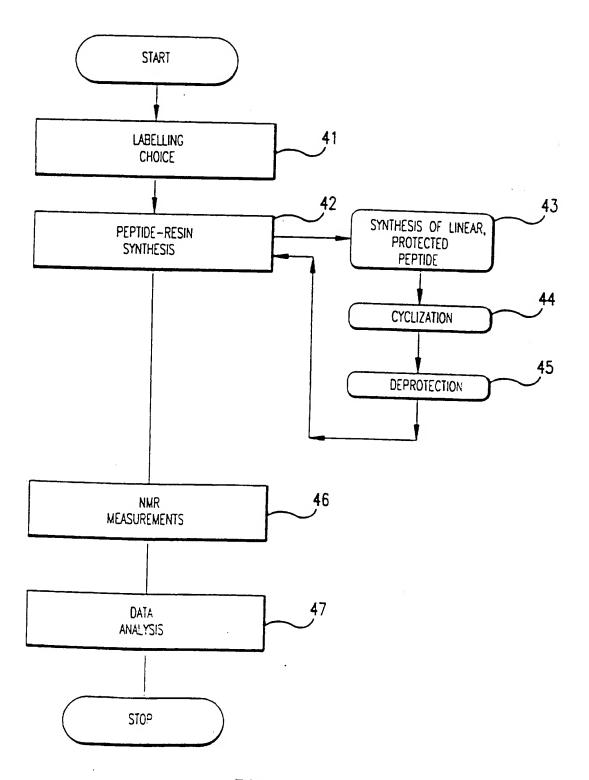
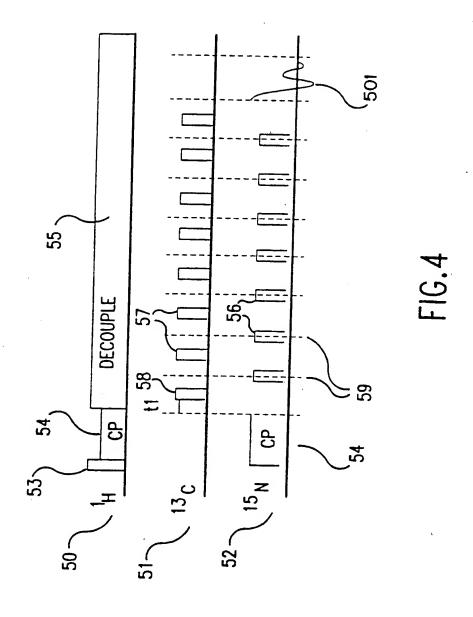
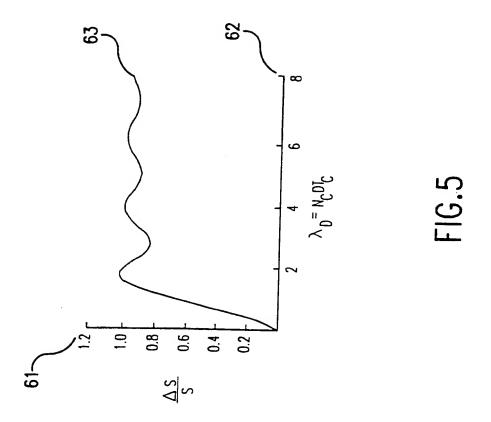


FIG. 3
SUBSTITUTE SHEET (RULE 26)

4/21



SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

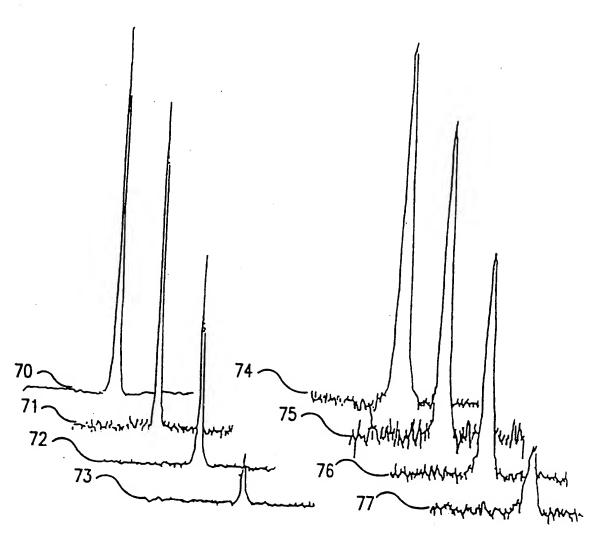
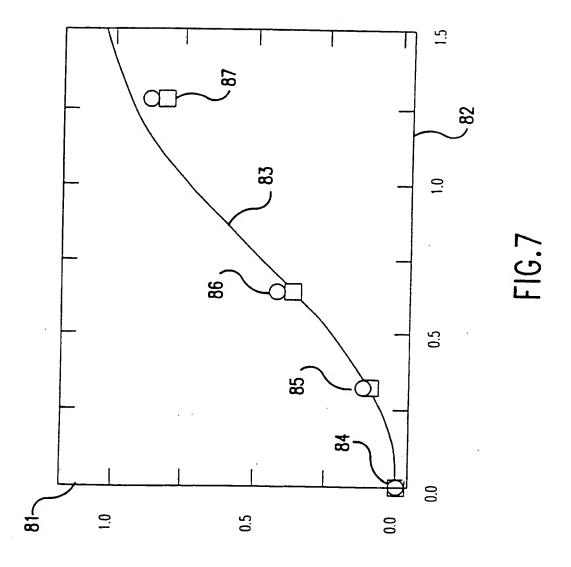


FIG.6

SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

8/21

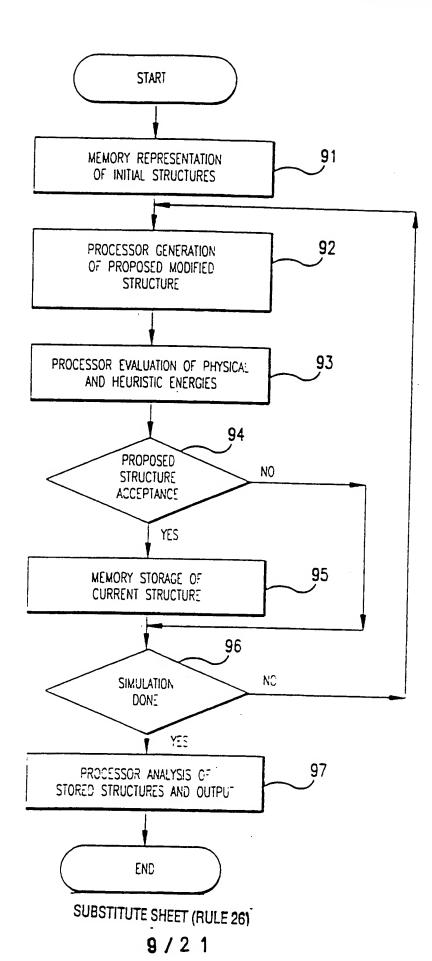
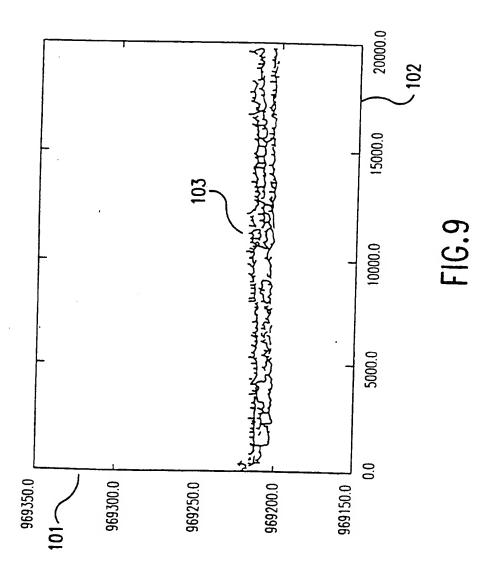
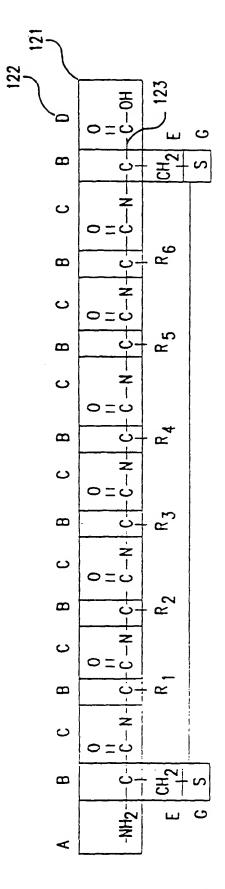


FIG.8



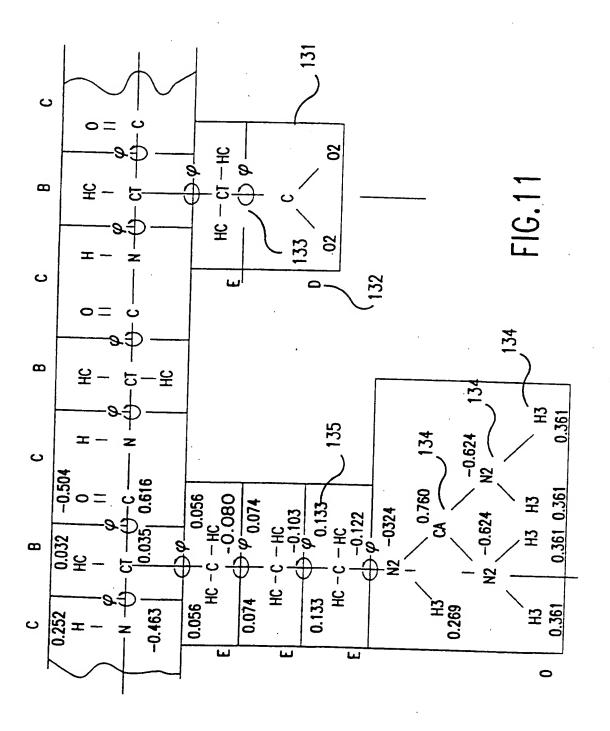
SUBSTITUTE SHEET (RULE 26)

10 / 2 1

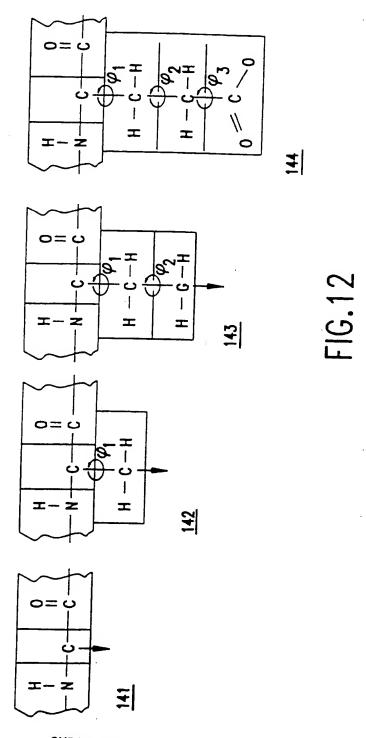


SUBSTITUTE SHEET (RULE 26)
11 / 2 1

FIG. 10



SUBSTITUTE SHEET (RULE 26)
12 / 2 1



SUBSTITUTE SHEET (RULE 26)

13 / 2 1

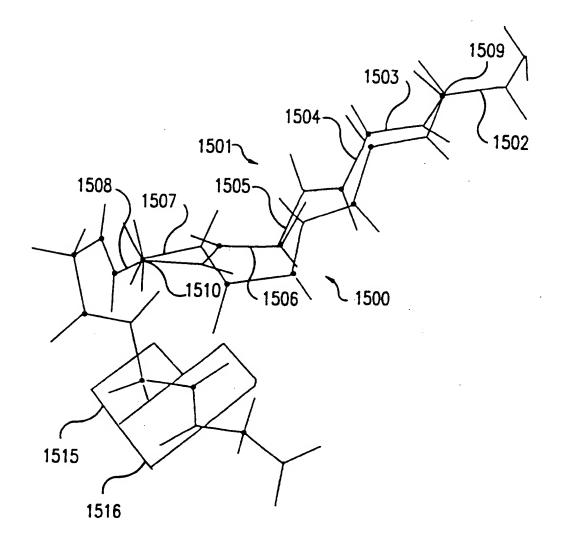


FIG.13

SUBSTITUTE SHEET (RULE 26)

14/21

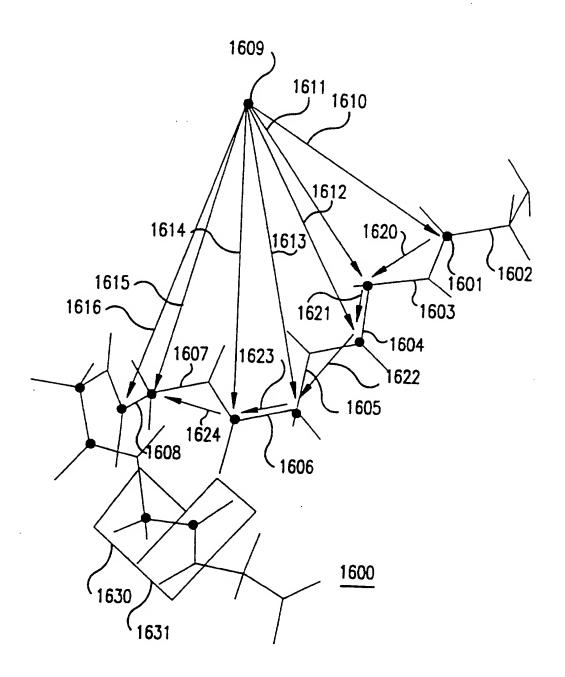


FIG. 14
SUBSTITUTE SHEET (RULE 26)
15 / 2 1

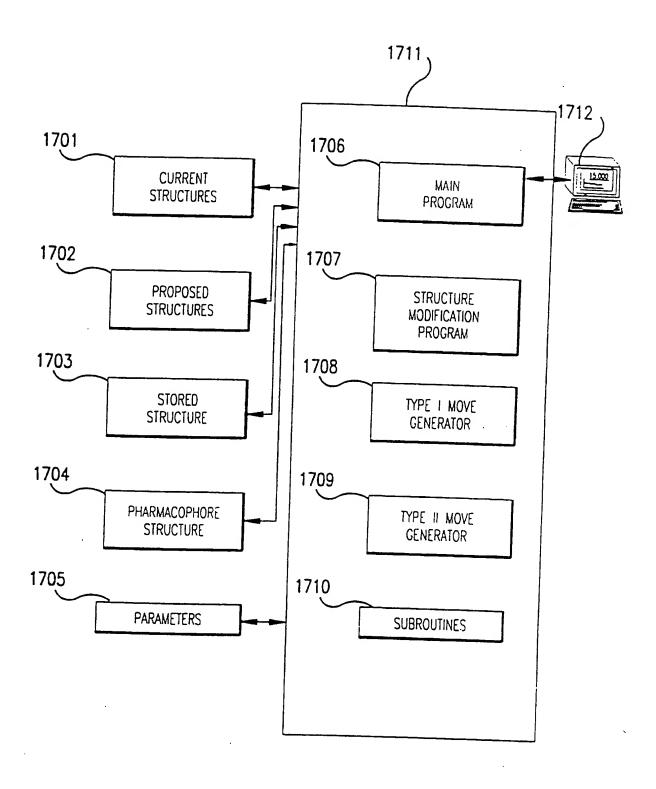
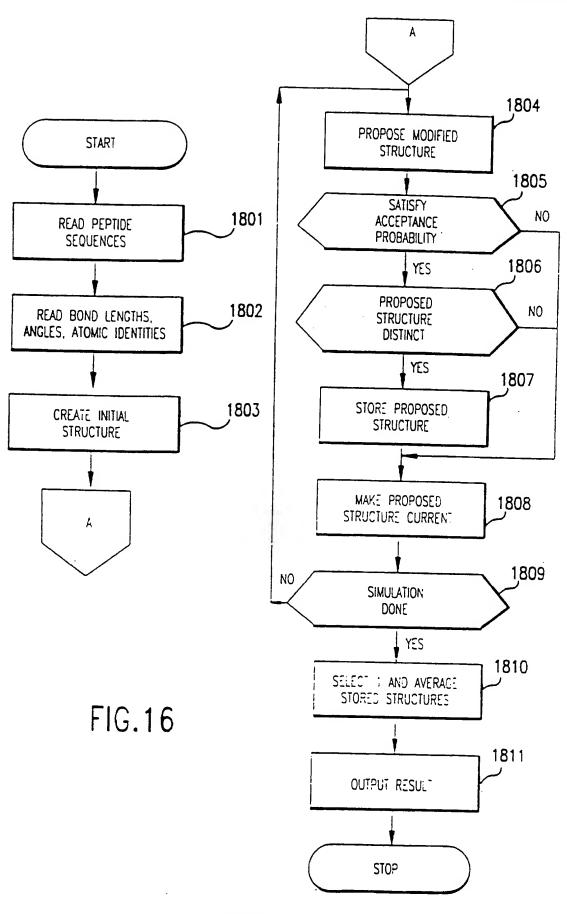


FIG. 15
SUBSTITUTE SHEET (RULE 26)
16 / 2 1



SUBSTITUTE SHEET (RULE 26)
17 / 2 1

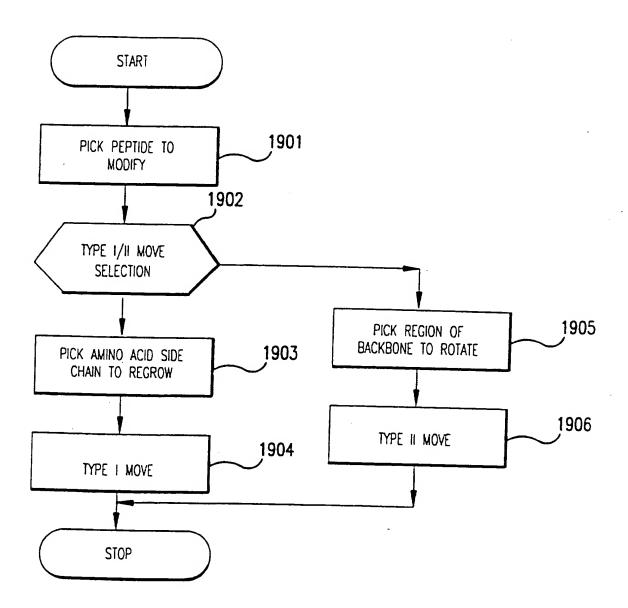
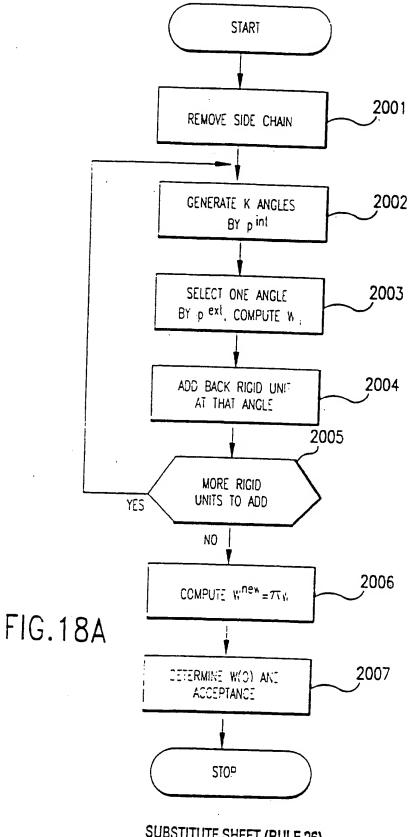
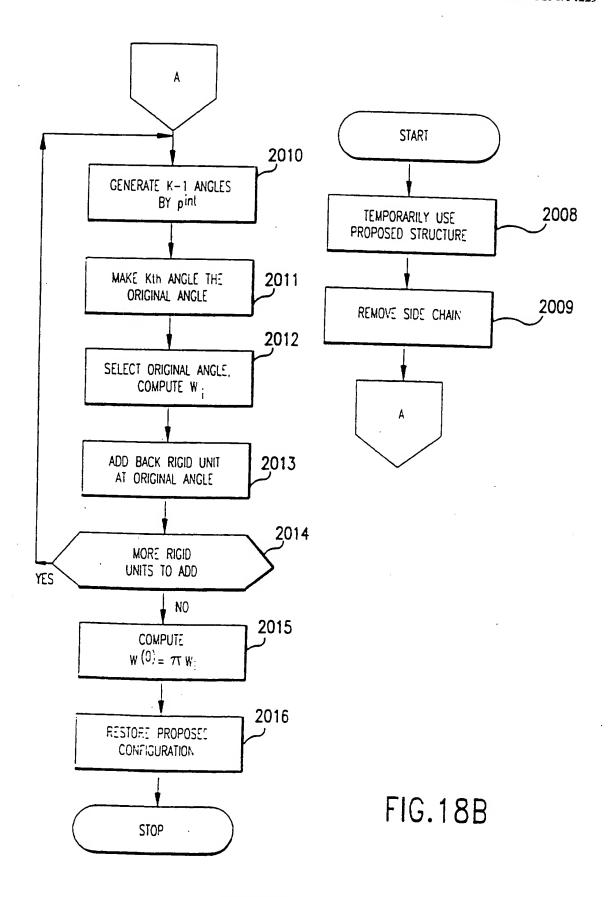


FIG. 17
SUBSTITUTE SHEET (RULE 26)
18 / 2 1

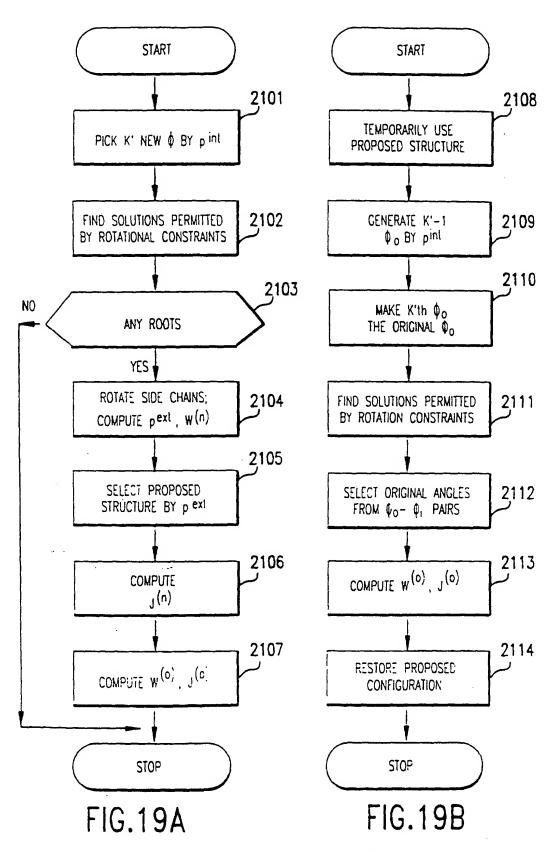


SUBSTITUTE SHEET (RULE 26)
19 / 2 1



SUBSTITUTE SHEET (RULE 26)

20 / 2 1



SUBSTITUTE SHEET (RULE 26)
21 / 2 1

International application No. PCT/US96/04229

IPC(6) US CL	SSIFICATION OF SUBJECT MATTER :G06F 17/50, 159:00; G06G 7/58; G01N 24/12, 3: :364/413.01, 496, 578; 436/173, 501, 518; 435/7.1 to International Patent Classification (IPC) or to bot			
B. FIE	LDS SEARCHED			
Minimum c	ocumentation searched (classification system follow	ed by classification symbols)		
U.S. :	364/413.01, 496, 578; 436/173, 501, 518; 435/7.1	·		
Documenta	tion searched other than minimum documentation to the	he extent that such documents are included	I in the fields searched	
APS, ST	lata base consulted during the international search (r N erms: library, pharmacophore, REDOR NMR, m	·	, search terms used)	
c. Doc	UMENTS CONSIDERED TO BE RELEVANT			
Category*	Citation of document, with indication, where a	ppropriate, of the relevant passages	Relevant to claim No.	
X A	HODGKIN et al. A Monte Carlo Pharmacophore Generation Procedure: Application to the Human PAF Receptor. Journal of Computer Aided Molecular Design. 1993, Vol. 7, pages 515-534, see pages 517-521.			
			1-21, 25, 28- 41, 108	
×	WILSON et al. The Calculation and Synthesis of a Template		51-106	
A	3663, see entire document.	TOV et al. Library of libraries: Approach to synthetic binatorial library design and screening of macophore" motifs. Proceedings of the National emy of Sciences. June 1995, Vol. 92, pages 5426-		
P, A	combinatorial library desig "pharmacophore" motifs. Proce			
X Furth	er documents are listed in the continuation of Box C	See patent family annex.		
* Special categories of cited documents T* later document published after the international filing date or priority date and not in conflict with the application but cated to understand the principle or theory underlying the invention.			iling but cited to understand the entirin	
E carbor document published on or after the international filing date. N document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone.				
15 15*	cited to establish the publication date of another citation or other special (casion cas specified). *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.			
the	ament published prior to the international filing date but later than priority date channed	"&" document member of the same patent		
Date of the actual completion of the international search 03 JULY 1996		Date of mailing of the international search report 25 JUL 1996		
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. (2023)		Authorized officer LUGUA FAUR 1 LORA M. GREEN		
	703) 305-3230	Telephone No. (703) 308-0196		

International application No. PCT/US96/04229

C (Continue	ation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages		Relevant to claim No.
A	BIANCHI et al. A Conformationally Homogenous Combinatorial Peptide Library. Journal of Molecular Biology. 1995, Vol. 247, pages 154-160, see abstract.		1-41, 108
X	US 5,265,030 (SKOLNICK ET AL.) 23 November 1993, see column 2, line 21-column 3, line 20.		68-102, 104-106
Y	GARBOW et al. Determination of the Molecular Confirmation of Melanostatin Using 13C,15N-REDOR NMR Spectroscopy. Journal of the American Chemical Society. 1993, Vol. 115, pages 238-244, see Experimental Section.		42-50, 107
Y	FERNANDEZ et al. Magnetic Resonance Studies of Polypeptides Adsorbed on Silica and Hydroxyapatite Surfaces. Journal of the American Chemical Society. 1992, Vol. 114, pages 9634-9642, see abstract.		42-50, 107
	· ·		
		ļ	
Ì			•

International application No. PCT/US96/04229

Box I ()hservations where certain claims were found unsearchable (Continuation of item I of first sheet)				
This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:				
Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:				
2. Claims Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:				
Claims Nos.: hecause they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).				
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)				
This International Searching Authority found multiple inventions in this international application, as follows:				
Please See Extra Sheet.				
As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.				
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.				
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:				
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.;				
Remark on Protest The additional search fees were accompanied by the applicant's protest.				
No protest accompanied the payment of additional search fees.				

International application No. PCT/US96/04229

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-41 and 108, drawn to a method of determining a consensus pharmacophore structure.

Group II, claim(s) 42-50 and 107, drawn to a method of making solid state magnetic resonance methods.

Group III, claim(s) 51-67 and 103, drawn to a method of configurational Monte Carlo determination.

Group IV, claims 68-102 and 104-106, drawn to an apparatus for configurational bias Monte Carlo determination.

The inventions listed as Groups I-IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Groups I and Groups II and III do not share a special technical feature as each has different steps and different end results. The method of Group I requires screening of a diversity library and generation of a pharmacophore, neither of which is required by the methods of Groups II and III. The distances required for the pharmacophore generation of Group I could be obtained by methods other than through the use of solid-state NMR methods ofGroup II, such as by solution NMR or X-ray crystallography. In addition, the method of Group I as claimed does not require the method of Group III as claimed, as the dependant claim of Group I which specifies the Monte Carlo method (claim 24) requires generating a proposed structure based on data for diversity libraries, whereas the method of Group II does not require the use of diversity libraries. Thus, Groups II and III lack the special technical feature of Group I, i.e. using a diversity library to generate a consensus pharmacophore structure.

Groups II and III are related as separate methods, as Group II is drawn to a method of making NMR measurements, while the method of Group III is drawn to a method of configurational monte carlo analysis. Thus, Groups II and III do not share a technical feature.

Groups I and II are related to Group IV as separate methods and product, as the methods of Groups I and II as claimed do not require the apparatus of Group IV as elaimed.

Groups III and IV are related as separate method and product. The method of Group III as claimed does not require the use of the apparatus of Group IV as claimed. In addition, the apparatus of Group IV could be used in methods other than the method of Group III such as use generating structures using NMR data obtained from a compound in solution masse.